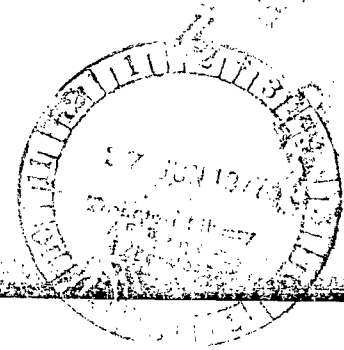
NASA
TT
F-794
c.1(R)

LOAN COPY: RETU
AFWL TECHNICAL I
KIRTLAND AFB,

(NASA-TT-F-794) PRINCIPLES OF TECHNICAL PLANNING FOR SATELLITE COMMUNICATIONS SYSTEMS (Kanner (Leo) Associates) 373 p HC $21.75 CSCL 17B

N74-34600

Unclas
G3/07 51192



TECH LIBRARY KAFB, NM
0069157

1. Report No. NASA TT F-794	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Principles of Technical Planning for Satellite Communications Systems		5. Report Date August 1974 6. Performing Organization Code
7. Author(s) A. D. Fortushenko, G. B. Askinazi, V. L. B kov, L. Ya. Kantor and O. S. Krapotin		8. Performing Organization Report No. 10. Work Unit No.
9. Performing Organization Name and Address Leo Kanner Associates P. O. Box 5187 Redwood City, California		11. Contract or Grant No. NASw-2481 13. Type of Report and Period Covered Translation
12. Sponsoring Agency Name and Address NASA, Code KSS-1 Washington, D. C. 20546		14. Sponsoring Agency Code

15. Supplementary Notes Translation of "Osnovy Tekhnicheskogo Proyektirovaniya Sistem Svyazi Cherez ISZ," Moscow, Svyaz' Press, 1970.

16. Abstract This book describes the basic properties of satellite communications systems, their significance and position among existing types of communications, studies organizational and technical principles of the construction of systems, problems of control of systems, problems related to the motion of satellites in orbit, and frequency ranges suitable for radio communications with satellites. Recommendations are presented for the performance factors and their basis is noted. A complete method is presented for calculating the signal power requirements of a communications system and a method is developed for calculating the basic parameters of the high frequency links of a satellite communications line.

The book is designed for specialists in radio communications, and may also be of interest for teachers and students in electronics schools.

18 tables, 157 figures, 97 biblio. refs.

17. Key Words (Selected by Author(s))	18. Distribution Statement		
	Unclassified. Unlimited. STAR Category 07		
19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price *
None	None	374	$8.00

*For sale by the National Technical Information Service, Springfield, Virginia 22151

TABLE OF CONTENTS

Foreword	v
List of Symbols	ix
Chapter 1. Purpose and Functions of Satellite Communications Systems	1
Chapter 2. Design of a Satellite Communications Network with Unassigned Channels	15
Chapter 3. Design of Satellite Communications Networks	57
Chapter 4. Peculiarities of Satellite Communications Systems Resulting from the Movement of the Satellite Relay Station	102
Chapter 5. Frequency Ranges Used for Communications by Satellite and Peculiarities of Their Use	138
Chapter 6. Performance Factors of Satellite Communications Channels	157
Chapter 7. The Energetics of Satellite Communications Lines	169
Chapter 8. Multiple Access Operation and Methods of Multiplexing of Radio Signals	234
Chapter 9. Parameters of the FM Transmission System of a Communications Satellite Link	290
Chapter 10. Mutual Interference Between Satellite Communications Systems and other Systems	322
Chapter 11. Principles of Construction of Control Apparatus in Satellite Communications Networks	335

PRECEDING PAGE BLANK NOT FILMED

The new aspects of the problem of satellite communications were reflected in the stream of journal articles which appeared, in which particular problems were solved. More fundamental works encompassing the entire problem have essentially never been published.

In presenting the present monograph to the reader, the authors /4 have attempted to create a rather complete text on the planning of communications satellite systems. In it, we studied problems of a purely systems nature, discussing the principles of construction and functioning of satellite systems. At the present time, the authors are preparing a monograph on the planning of satellite communications equipment, which will be a logical continuation of the present book. The new book will describe the hardware decisions influencing the parameters of satellite systems. In the text of this book, we refer to the book you hold in your hands as the first volume, to the book currently in preparation as the second volume.

A number of planning problems have been studied practically for the first time: these include problems of the organization and control of complex communications satellite networks, calculation of their throughput capacity, the problem of selection of a method of modulation to support the operation of the entire network through a relay satellite, and evaluation of economic effectiveness.

Some sections contain a planning method allowing timely and rather precise calculation of the parameters of the individual links and the communications system as a whole. These include the sections on determination of performance factors, transmission power levels, mutual interference between ground and satellite systems and the selection of parameters of the microwave link. Finally, some chapters discuss hardware problems in the planning of Earth station and satellite equipment, to the extent that they influence the parameters and planning of the entire system. Description of the physical principles of satellite communications and the well-known problems of the operation of communications links is presented in abbreviated form, as necessary to maintain logical sequence and orderliness in the presentation.

The range of problems reflected in the book is quite broad. However, the reader will not always find exhaustive answers to the questions of interest to him, since a number of the problems have not yet been solved. Nevertheless, the authors hope that the book will be useful for specialists with sufficient training in the area of electronics, involved in the development and creation of satellite communications systems. It will be of interest for graduate students and students in the upper classes of colleges specializing in space communications.

The writing of the first volume was distributed among the authors as follows: Chapter 1 was written by Doctor of Technical

FOREWORD

/3*

The problem of the transmission of information using extra-terrestrial objects and, in particular, artificial Earth satellites, has arisen quite recently. Its practical aspects began developing in 1957 after the launch of the first artificial satellite in the USSR. Due to its pressing nature and complexity, this problem immediately attracted the attention of a broad range of scientists.

The successful launch of the first communications satellites in 1960-1962, providing the basis for the creation of new long-range radio communications links, and the initial experience of the operation of these links served as the basis and stimulus for rapid development of theory and practice of satellite radio communications. Many books and journal articles appeared, both in the USSR and abroad, setting forth the physical capabilities, basic peculiarities and even methods of engineering design of communications lines using satellites. This initial stage is characterized primarily by an interest in the principles of the new type of communications, as applicable to links consisting of two Earth stations and one relay satellite.

The next stage in development, beginning with the launching of the Molniya-1 and Early Bird satellites in 1965, allowed a transition to be made from the construction of links to the construction of satellite communications systems, including large numbers of Earth stations and satellites. This transition to the new stage in the development of satellite communications in the USSR was marked by the creation of the Orbita multi-station network, with over 20 Earth stations, in 1967. Regular transmission of television signals was achieved with the Molniya-1 satellites, rotating in high elliptical orbits.

We should note that the second stage was new not only in its quantitative aspect, but in the qualitative aspect as well. Consideration of the combined operation of several satellites and a large number of Earth stations, problems of their interaction and, finally, the natural desire to make these systems operationally convenient and economical, has required a new approach to the planning of systems, since the technical decisions considered optimal for communications satellite links were frequently unsuitable for systems. The corresponding changes were made to apparatus as well, particularly the equipment carried on the satellites, the range of requirements forwhich expanded significantly.

---

*Numbers in the margin indicate pagination in the foreign text.

Sciences, Professor A. D. Fortushenko, Chapters 2, 3 and 11 by Candidate of Technical Sciences G. B. Askinazi, Chapters 7 and 8 by Candidate of Technical Sciences V. L. Bykov, Chapters 6 and 9 by Candidate of Technical Sciences L. Ya. Kantor, Cahpters 4, 5 and 10 by Engineer O. S. Krapotin. A number of sections of Chapter 8 were written by Graduate Student V. A. Borovkov.

All critical comments and suggestions should be sent to Svyaz' Press (Moscow-Center, Chistoprudnyy Boulevard, 2).

****

NASA Translation Editor's Note: The NASA translation was prepared with the consultation of Mr. John M. Barstow and with support from the COMSAT Corporation.

****

I

PRECEDING PAGE BLANK NOT FILMED

LIST OF SYMBOLS /5

A -- azimuth:
A -- non-uniformity of frequency characteristics of HF link;
a -- amplitude of compressed radio signals;
a -- semimajor axis of elliptical orbit;
b -- semiminor axis of elliptical orbit;
B -- equipment utilization effectiveness factor;
b -- gain in signal/noise ratio provided by receiver in general;
B(FM) -- gain in signal/noise ratio provided by FM receiver;
B_n(FM) -- same, for noise of another type;
C -- throughput capacity of communications channel;
c -- speed of light;
D_a -- diameter of antenna;
d -- detection parameter;
d -- distance between satellite and ground station;
E -- signal energy,
e -- eccentricity of elliptical orbit;
e_1, e_2 -- ellipticity factors of antenna polarization;
F -- shielding factor of terrain at ground station;
F -- oscillating frequency in video channel;
F_m -- modulating frequency;
F_B -- highest frequency of video spectrum;
ΔF_k -- band of frequencies of telephone channel;
ΔF_{gr} -- width of band of group spectrum;
f -- frequency in radio link;
f_0 -- carrier frequency;
f_{dop} -- frequency, changed as a result of the Doppler effect;
f_k -- middle frequency of upper telephone channel;
Δf -- band of frequencies of receiver;
Δf -- band of frequencies occupied by signal of one stage;
Δf_g -- guard band between signals of neighboring stations;
Δf_n -- equivalent noise band of receiver;
Δf_d -- peak deviation of frequency of multichannel message;
Δf_{dk} -- peak deviation of frequency per telephone channel;
$\Delta f_{dk\ eff}$ -- effective deviation of frequency per telephone channel;
Δf_{dop} -- Doppler shift of carrier frequency;
G -- gain of antenna relative to isotropic radiator; /6
G_s -- gain of on-board satellite antenna;
G_g -- gain of ground station antenna;

PRECEDING PAGE BLANK NOT FILMED

G_r -- gain of receiving station antenna;

G_{tr} -- gain of transmitting station antenna;

G_i -- gain of transmitting antenna of interfering station in the direction of the receiving station experiencing noise;

H -- altitude of satellite over the surface of the Earth;

H_a -- altitude at apogee;

H_p -- altitude at perigee;

H(x) -- entropy of random quantity x;

i -- inclination of orbit;

J -- Bessel function;

K -- usage factor of channels in communications network;

K -- constant of field of gravity of Earth;

K_s -- suppression factor of weak signal by strong signal;

k_p -- psophometric factor;

K -- channel load factor;

k -- Boltzman's constant,

k_2, k_3 -- factors of harmonics (second and third);

k_{AM-PM} -- AM-PM conversion factor;

L_o -- primary losses of signal during propagation;

L'_o -- same losses, adjusted to frequency of 4 GHz;

L_{add} -- additional losses during propagation;

L_Σ -- summary losses between transmitter and receiver;

l_s -- number of channels in communications net;

l_d -- number of channels in communications net;

M -- mass of the Earth;

m -- mass of the satellite;

m -- frequency modulation index;

m -- number of communications lines emitted from a station;

N -- total number of telephone channels transmitted in relay trunk;

N_s -- number of simplex channels of relay station;

N_d -- number of duplex channels of relay station;

N' -- number of channels transmitted by one station;

n -- number of signals multiplexed;

n -- number of subtrunks in a network;

P -- power;

P_s -- signal power;

P_n -- power of noise at input of receiver;

P_s -- power of satellite (transmitter);

P_g -- power of ground station transmitter;

P_{gi} -- power of Earth station interfering transmitter;
P_n -- power of noise;
$P_{n\Sigma}$ -- power of all components of noise at input of demodulator;
P_{nk} -- power of all components of noise in telephone channel;
P_{nt} -- power of thermal noise in channel;
P_{nn} -- power of nonlinear noise; /7
P_{nt} -- power of transient noise;
$P_{a\ AM-PM}$ -- power of intelligible interference of AM-PM conversion;
$P_{i\ AM-PM}$ -- power of unintelligible interference of AM-PM conversion;
P_{nk} -- power of noise in telephone channel;
P_i -- power of signal of one station;
p -- losses (probability of failure of incoming call);
p -- probability;
p -- focal parameter of elliptical orbit;
Q_g -- capacity of ground station (in channel);
Q_i -- number of channels in ith ground station;
Q_Σ -- capacity of entire network (in channels);
Q_r -- capactiy of relay station;
Q_a -- volume of actively used equipment;
Q_p -- volume of equipment potentially available;
Q_{out} -- number of channels of outgoing communications;
Q_{in} -- number of channels of incoming communications;
q -- antenna surface utilization factor;
q_f -- frequency duty factor;
q_t -- time duty factor;
R_e -- radius of the Earth;
r -- number of channels in a bundle, subtrunk or link;
r -- radius-vector of satellite (distance of satellite from center of Earth);
r_a -- apogee distance;
r_p -- perigee distance;
S -- number of stations in communications network;
S_A -- geometric area of antenna aperture;
$S(\omega)$ -- signal spectrum;
s -- number of revolutions of satellite around the Earth per day;
T -- time of one revolution of satellite;
T -- cycle interval;
T_s -- length of a sidereal day;

T_A -- equivalent antenna noise temperature;
$T_{\Sigma b}$, $T_{\Sigma g}$ -- full equivalent noise temperature of on-board and ground station receivers;
t_m -- mean time required to service one call;
t_0 -- moment when satellite passes perigee;
U -- amplitude of signal;
V -- velocity of movement of satellite in orbit;
V_{sup} -- velocity of motion in direction away from ground station;
v -- true anomaly;
W -- band of frequencies of relay station;
W_i -- band of frequencies of partial channel;
Y -- stream of messages (in erlangs);
Y_{out} -- outgoing stream;
Y_{in} -- incoming stream;
y -- load (in call-hours);
Z -- effectiveness factor of radio compression method;
α -- angular radius of zone of visibility relative to center of Earth; /8
β -- elevation angle of ground station antenna;
γ -- time compression factor;
γ -- angular error of orientation of on-board antenna toward center of Earth;
γ -- angular radius of zone of visibility relative to satellite;
Δ_{nt} -- gain with respect to thermal noise resulting from introduction of predistortion;
Δ_{2n}, Δ_{3n} -- gain with respect to nonlinear second and third order noise resulting from introduction of predistortion;
θ -- angle of inclination of velocity vector of satellite to plane perpendicular to radius vector;
θ_0 -- maximum number of calls arriving per second in a fixed stream with fixed losses;
θ_1 -- geographic longitude;
ν -- peak factor of multistation signal;
ν_i -- peak factor of signal of one station;
ν_N -- peak factor of N-channel group telephone signal;
τ -- length of signal of one station;
τ_g -- guard time interval between signals of ground stations;
τ_0 -- time of processing one request at control center;
Ω -- longitude of ascending node;
Ω -- modulating frequency;
ω -- argument of perigee;
ω -- frequency;
κ -- load concentration factor;

λ -- wavelength;
λ -- intensity of stream of requests;
ρ -- intensity of servicing;
σ^2 -- spectral density of noise at output of receiver;
σ_0^2 -- same at input of receiver;
ϕ -- phase;
ϕ_1 -- geographic latitude;
$\Delta\phi_{0.5}$ -- width of antenna radiation pattern at half-power level;
$\Delta\phi$ -- phase modulation index;
η -- efficiency;
η_b, η_g -- efficiency of wave channel of on-board and ground-based transmitters.

PRINCIPLES OF TECHNICAL PLANNING FOR SATELLITE COMMUNICATIONS SYSTEMS

A. D. Fortushenko, G. B. Askinazi, V. L. Bukov
L. Ya. Kantor and O. S. Krapotin

Chapter 1. Purpose and Functions of Satellite Communications Systems /9

§ 1.1. Principles and Basic Peculiarities of Satellite Communications Systems

The principle of communications by satellite consists in transmission of signals from one or more ground stations to an Earth satellite and subsequent relay of the signals to all ground stations of the system (Figure 1.1). This allows each ground station of the network to communicate with any other station of the same network. If the corresponding stations observe the satellite simultaneously, instantaneous relay is possible; if the stations are not both within the zone of simultaneous visibility of the satellite, in order for the information to be relayed it must be recorded on the satellite, so that relay involves a time delay (relay with memory).

Depending on the nature of the processing of the signals on-board the satellite, relay may be passive, active or quasi-passive. In the first case, the satellite transmits the electromagnetic oscillations by simple reflection without converting them or changing their structure. The advantages of passive relay include extreme simplicity of the satellite, high reliability, no mutual interference between the signals of ground stations and, as a result, unlimited throughput capacity of the passive relay station. The shortcomings include low power detectors, resulting from the great losses of power during propagation of radio waves, requiring the use of extremely complex and cumbersome antennas, powerful transmitters and sensitive receivers at the ground stations.

In active relay, the functions of loss compensation are divided between the ground station equipment and the satellite which in this case receives, amplifies and when necessary processes (changes the structure of) the signal before radiating it back to the Earth. The complication of the satellite allows a significant reduction in the power requirements of the ground station. /10 However, the reliability of the on-board complex is reduced; it is therefore desirable to study the degree of complication of the satellite carefully in planning.

Active relay stations have limited frequency ranges and, consequently, can be used for transmission of only certain numbers of signals. In order to increase the capacity, additional relay equipment must be installed on board the satellite. However, the possibility of producing a gain in power is decisive and fully compensates for the shortcomings.

Quasipassive relay occupies an intermediate position between passive and active relay and is characterized by the presence of the minimum, primarily passive, equipment on board the satellite, performing some amplification of the signal as it is relayed by concentrating the power in the required direction. The peculiarities of quasipassive relay include increased secrecy and interference stability of the information transmitted.

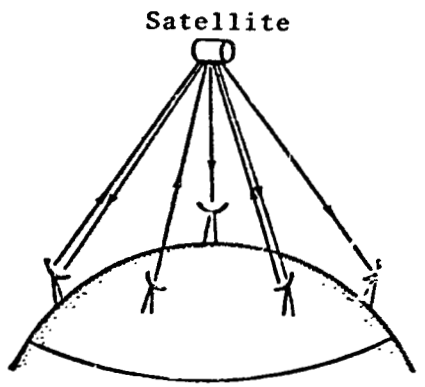


Figure 1.1. Principle of Satellite Communications

During the first stages of development of satellite communications, the question as to the selection of the type of relay was discussed broadly and comprehensively. At present, we can state with some assurance that it has been solved in favor of active relay, which is most commonly used and has almost completely replaced the other two types.

In accordance with this, further analysis will be confined primarily to systems with active relay and without memory.

This principle of transmission of information by satellite defines all the capabilities and peculiarities of satellite communications systems. The basic peculiarity of the systems which we will study, distinguishing them from ground systems, is that the relay station, installed on the satellite, can service a considerable area on the surface of the Earth simultaneously, the area increasing with increasing altitude of the satellite and reaching, in the limiting case, almost half of the surface of the Earth (Figure 1.2), whereas surface communications links operate in a given direction and can service only the adjacent regions[1]. /11

[1]The only exception is short wave radio communications, which can be used to provide communications in a zone; however, a short wave communications line has limited throughput capacity, low quality and low reliability, and therefore cannot compete with satellite communications systems.

Covering a significant territory, the relay satellite allows any two points in the area to communicate, i.e., allows communications to be provided in any direction. Creation of a communications line requires only the construction of ground stations with antennas directed toward the satellite and the necessary accessory equipment. This property makes satellite communications irreplaceable under complex geographic conditions, when great expanses of water, mountains, forests or swamp must be crossed, so that the construction of surface communications lines involves great organizational or construction difficulties.

One advantage of satellite communications is the possibility of simultaneous operation of a large number of ground stations through a single relay station. The communications may be either radial, i.e., one station (central) with all other stations, or "each with each," allowing simultaneous communications of all stations with each other.

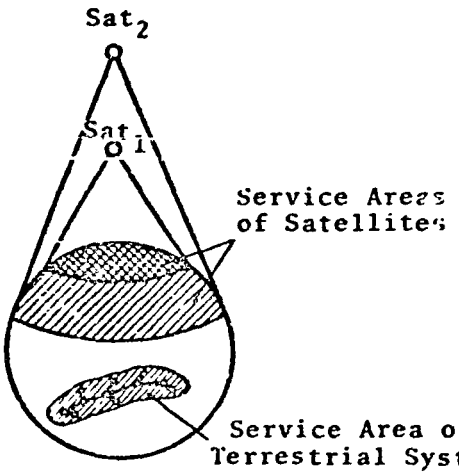


Figure 1.2. Comparative Evaluation of Areas Served by Communications Satellite and Surface Line

The creation of satellite communications systems involves great capital and operational expenditures, but the cost of satellite communications channels is independent of the distance between ground stations. This means that these systems are economically effective for communications between widely separated points on the Earth.

Satellite communications lines, like surface cable and radio relay lines, are universal with respect to the type of information carried; due to their great bandwidth and throughput capacity, they can be used for transmission of television and radio broadcast programs, transmission of multichannel telephone messages, telegraph and facsimile signals, as well as digital information. The performance factors of satellite communications channels can be maintained quite high, corresponding to the norms for surface lines.

§ 1.2. Functions of Satellite Communications Systems in the General Communications Network /12

One of the basic problems arising in evaluation of a new type of communications is determination of its role and place among the available communications equipment. The properness of the solution will determine the harmony of the interaction of various communications systems and, in the final analysis, their effectiveness.

Satellite communications systems are no exception. The unfounded opinion, commonly held recently, that satellite communications can replace all other types of communications, has now been replaced by a more mature evaluation, according to which satellites will have a more modest role, that of supplementing and backing up existing terrestrial communications systems.

This is understandable. Satellite communications lines are similar in their peculiarities to surface trunk lines, which are terminated by complex apparatus and are used for communications between remote communications centers or large telephone exchanges; further connection with subscribers must be made using surface lines.

Thus, satellite communications networks cannot be fully independent (autonomous). They form but a part of the overall network, encompassing both surface and satellite lines. The role of the satellite network in the overall communications network is determined by its primary properties -- flexibility, maneuverability and the ability to cover large areas.

The simplest, most obvious function of a satellite communications network is the transmission of broadcast type information: television, radio broadcasts, etc. In some cases, use of satellites for this purpose allows transmission of television to remote regions significantly more cheaply than surface lines.

Particularly great advantages of the use of satellites for the transmission of broadcast information are to be expected in those cases when this information must be transmitted to points at some distance from the center and from each other. On the other hand, when broadcast information is to be transmitted to a large number of nearby points, it is better to use surface lines, allowing transmission of programs over a single line connecting these points.

The second function of the systems in question is the organization of communications with inaccessible regions, when it is difficult and expensive to construct surface communications lines. The construction of one or more satellite ground

stations in these cases may be significantly less expensive than the construction of a surface communications line, or at times may be the only possible solution.

Finally, yet another extremely important function of satellite communications lines is that they can and should be used as a means for maneuvering within communications networks. Actually, overloads arise periodically in any communications network in one direction or another. These overloads are explained primarily by the unevenness of the telephone load and they also result from failures in certain areas of the network, etc. In all these cases, it is quite important to be able to vary the structure of the network, i.e., to increase the capacity of communications lines in the overloaded sector by including additional lines and creation of bypasses or other methods. In surface communications networks, maneuvering of capacities and information flows is very difficult, since this maneuvering requires that additional junctions and stations be activated, that information be transmitted over extremely long paths, frequently over poor lines. Satellite communications lines can be organized in the area required very rapidly and easily if, of course, the ground stations are in place. /13

Satellite communications lines can be used not only to increase the capacity of certain network sectors, in which random overloads occur; these lines can be effectively used for regular operation together with surface lines as well.

As we know, surface lines are fully loaded only during the hours of maximum load, representing a small fraction of the day, and the rest of the time are not utilized to more than 30% of capacity. If the lines of the surface network are constructed to handle the mean daily load, and the necessary number of additional channels is provided during the peak hours by means of satellite communications, the overall effectiveness of the operation of the combined network will be significantly higher, particularly if we consider that the hours of maximum load in different directions are generally shifted relative to each other in time and that the same satellite communications channels can be used for several such directions in sequence.

The interaction of surface and satellite networks determines their method of operation. The fact is that modern automated communications networks are usually designed to operate with cancellations and allow only a limited waiting time for establishment of a connection. The devices which control the waiting time are installed in switching junctions or telephone exchanges, i.e., on points on the path preceding the beginning of the satellite communications link

The satellite network has an independent automatic control and switching system. However, since waiting over the entire

network cannot be too great, it is not a good idea to introduce an additional waiting time to the satellite system, and the system must be designed to operate with failures, without waiting. The introduction of this limitation, as we will show below, predetermines the methods of control and calculation of the throughput capacity of the satellite network to a significant extent.

We must indicate a number of aspects of the use of satellite communications, particularly for communications with ships, aircraft and other moving objects. /14 Actually, any moving object in the zone of visibility of a satellite, if it is equipped with the proper apparatus, can transmit and receive signals to and from the satellite, like stationary ground points. The speed of movement of the object (even an aircraft) is no obstacle to the establishment of two-way communications.

Similarly, communications can be established between a satellite and a space probe or between two satellites. The latter is necessary when the Earth stations communicating with each other, due to their great separation, do not fall simultaneously in the zone of visibility of one satellite, and memory devices cannot be used on board the satellite for some reason (for example, due to their large volume).

Thus, with proper placement of ground stations and optimal utilization of the throughput capacity of the relay station, a satellite communications system can serve as a significant addition to the available surface lines and, in some cases, may be practically irreplaceable.

§ 1.3. Types of Systems. Characteristics of Territories Serviced

Satellite communications systems have great capabilities and may significantly influence the development and level of communications equipment.

According to their purpose, they may be divided into commercial and special systems. The former are designed for the transmission of civil and commercial information: television signals, multichannel telephone messages, telegraph, facsimile, etc. The latter are used for transmission of digital codes designed for communications with certain types of moving surface and nonsurface objects. Each of these types has its own characteristics and specific features. Here and in the following we will study only systems of the first type, i.e., systems designed for transmission of civil and commercial traffic.

As to area of coverage, more precisely size and territorial characteristics of area serviced, satellite communications systems

can be divided into local and global systems. In the former case, the system is used for the interests of one or several closely located countries and is used to satisfy the requirements for information exchange in the local area or with certain remote points. An example of this type of system is the system created by France and West Germany in cooperation [1.1]; Canada [1.2] is also creating a local system, to be used for communications between the east and west coasts in the country.

Local systems also include the Soviet satellite communications system based on the Orbita ground stations and Molniya-1 satellites. /15

Global systems are designed for exchange of information between any points on the Earth and, consequently, must include several satellites, the zones of visibility of which encompass all (or almost all) of the surface of the Earth. One well-known example of such a system consists of three satellites in stationary orbit, spaced at angles of 120° (Figure 1.3).

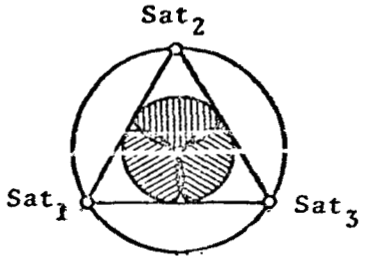


Figure 1.3. Global Satellite Communications System Using Stationary Satellites

More realistic examples include the Intelsat system [1.3] developed by the USA and the international system created by the USSR in cooperation with the socialist camp countries [1.4]. Although the type of system and its purpose are usually fixed in planning, it seems nevertheless useful to describe briefly the areas and territories covered by satellite communications.

The initial and basic criterion in selecting areas is the population density.

When local communications systems are planned, it is not difficult to draw up an overall picture of the distribution of population density over the territories of the participating countries. However, one must consider not only the existing distribution, but trends in its development (resulting in population growth or displacement), which makes the problem more difficult. The concentration of industrial and scientific centers is also significant, since they may be sources and consumers of large quantities of information.

The next stage in planning is determination of the traffic, i.e., the volume and direction of information transmission. Here, one should also consider the prospects for growth, since the

creation of a satellite communications system will involve at least three to five years of work. Further, the extent to which the traffic (and the prospective traffic) can be satisfied by existing and planned surface lines must be considered. The results produced allow rather precise establishment not only of the locations of ground stations, but also of the required throughput capacity of the satellite communications system planned, and the number of telephone channels required.

The number of television programs to be transmitted and the locations of receiving points can also be easily determined, using data on the distribution of population density and the plans of placement of television broadcasting equipment and program feed channels. One should consider here the time differences between territories in different time zones, and the various national (linguistic) compositions of the population. Both of these factors may be important only if the satellite television broadcast system covers countries with large territories and multinational populations. /16

Since the conditions which must be considered in the creation of local systems are specific for each case, it is impossible to give more concrete recommendations.

The planning of global communications systems should also be based on determination of the concentration and density of the population. However, great population density is not in itself an indication of great requirement for information exchange, since this requirement depends on the cultural level and commercial activity of the population. A more reliable characteristic is the distribution of telephone equipment, which indicates the presence and quantity of information directly.

Figure 1.4 shows the existing distribution of population density and telephones over the Earth as of 1960-1962, including the USSR and China [1.5]. We can see from Figure 1.4 that the overwhelming majority of telephone equipment is concentrated between 10° and 60° n. lat.; consequently, coverage of this territory in 1962 would have completely solved the problem of creation of a global satellite communications system.

Figure 1.5 shows the daily variations in requirements for telephone channels of all countries, excluding the USSR and China [1.5]. Using this figure, it is not difficult to establish that the existing requirements for exchange at that time could be satisfied with a throughput capacity of the satellite communications system of about 1000 duplex telephone circuits.

If we consider even the immediate prospects for development of communications, the results of the estimation are qualitatively and quantitatively different. Figure 1.6 shows a map of the /17

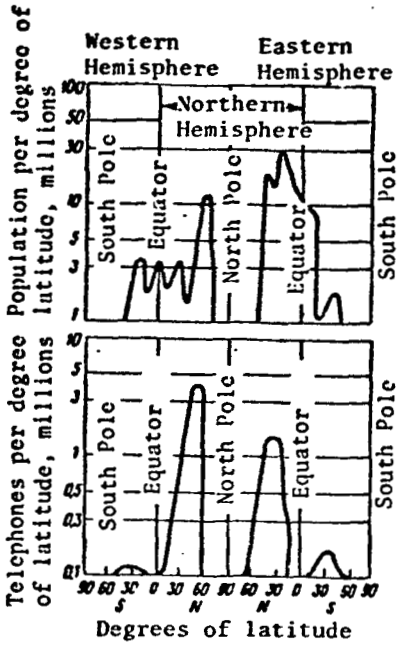


Figure 1.4. Distribution of Population Density and Telephones over the Earth

distribution of the main areas requiring international communications, considering the requirements of 1970 [1.6]. The economic and cultural development of the countries of Africa and Latin America have led to a significant expansion of the territory requiring coverage by satellite communications, a significant portion of this territory being in the southern hemisphere between 0 and 30° s. lat. The volume of information transmitted has also increased. According to the data of [1.6] for 1970, it has reached 20,000 duplex telephone channels and several tens of thousands of narrow band information channels. The number of duplex telephone circuits to be provided by satellite communications is approximately 12,000.

No complete and reliable information is available concerning the daily distribution of load in various directions, but we know that the most intensive intercontinental information exchange over the Europe-America lines is observed between 8 a.m. and midnight Greenwich Mean Time, with the American load almost double the Europe.n load.

From midnight until 8 a.m. Greenwich Mean Time, the maximum flow of information to Europe is from Australia. As concerns conversations between the antipodes, since the business hours of two locations do not correspond, the best time for conversations is when the waking hours overlap. For example, 9:00-10:00 in London and 21:00-22:00 in New Zealand.

Detailed consideration of the distribution of the daily load /18 allows proper calculation of the throughput capacity of a global system; however, this practical consideration is quite difficult due to the unavailability of statistical data.

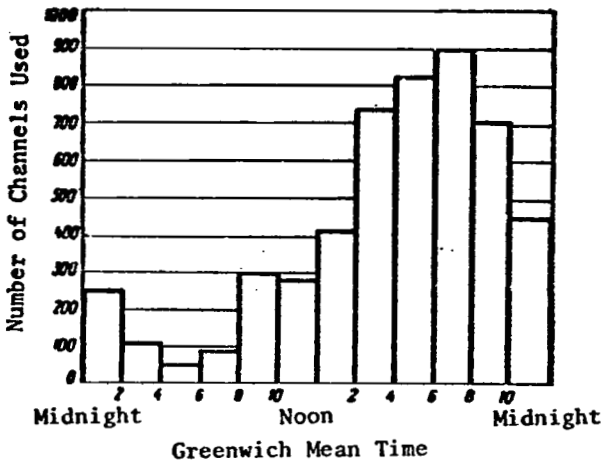


Figure 1.5. Daily Changes in Requirements for Telephone Circuits

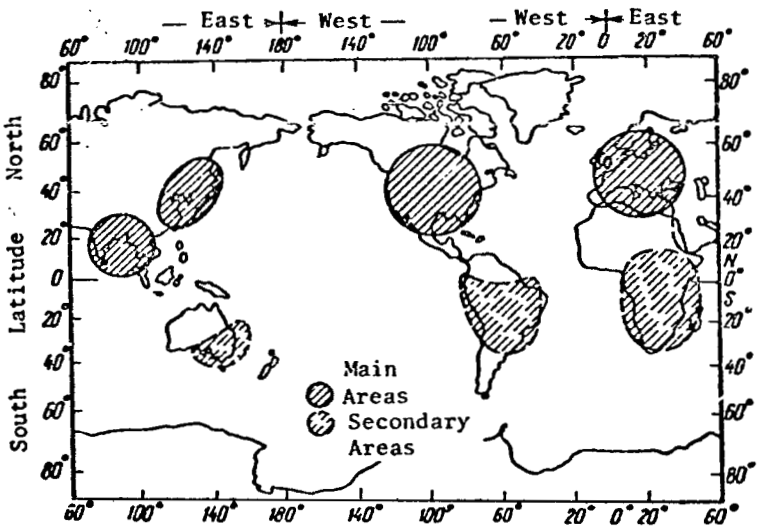


Figure 1.6. Distribution of Main Areas of International Communications

§ 1.4. Composition of a Satellite Communications System

A satellite communications system is an extremely complex system, consisting of a large number of different subsystems and individual units. Although this book studies essentially only the organizational side of a communications complex, nevertheless we must briefly determine the composition of the entire system in order to provide a proper presentation of the volume, role and interactions of the remaining complexes.

The satellite communications system includes:

1. The control center (CC) which performs technical administration of the operation of the entire system, including control of the operation of the ground stations and relay satellites and testing of their operating condition; here also are orders for lines and channels received and processed, statistical data produced on the operation of the system and satellite launchings to assure normal functioning planned.

2. Artificial Earth satellites equipped with a set of relay apparatus, with subsystems for electric power supply, orientation, control, temperature regulation and other service subsystems assuring normal functioning of the relay station during the fixed time of active existence of the satellite. /19

3. A rocket and space complex, combining the boosters and surface technological and launch equipment designed to place relay satellites in fixed orbital points.

4. Ground stations (GS) for satellite communications, equipped with transmitting and receiving equipment and designed for exchange of information of all types. The ground stations are divided into central, peripheral and mobile stations, depending on the volume of information transmitted and the function which they perform in the overall network of stations.

The central station (CS) occupies a leading position in the network, receiving and transmitting all types of information and coordinating the operation of all stations of the network. In addition to the main CS, there may be one or two reserve CS, the reserve CS acting as peripheral stations (PS) under normal conditions, making up the basis of the communications network. They exchange information between various points in the territory serviced. Mobile stations, in spite of their large number, are usually secondary in significance and are used for the transmission of small volumes of information, primarily special purpose information.

5. A test and control system, designed to perform measurements and determine the orbital parameters necessary to

predict the motion of satellites, to determine and send out data for calculation of instructions for all types of ground stations for aiming of their antennas toward the satellite, to switch equipment on the relay satellite on instructions from the ground and also to switch equipment to receive, record and process telemetry information from the satellite concerning the functioning of the on-board apparatus as it operates.

6. The system which guides antennas toward the satellite and tracks it with the necessary accuracy during each communications session.

7. Multiplexing apparatus, which combines telephone, telegraph, facsimile and other signals into common group signals which are then transmitted through the satellite communications link.

8. Surface connecting lines from ground stations to surface points (communications centers and television centers) which are the sources or consumers of the information transmitted.

9. Signal processing apparatus (apparatus for conversion of television signals, apparatus for regeneration of synchronization signals, echo-protective equipment, etc.), designed to match the structure of signals to the surface and space communications lines.

10. The communications automatic control system, designed to control channels and directions of communications for the purpose of more effective utilization of the throughput capacity of the relay station and optimization of the operation of the entire network. /20

11. A service communications system, used to bring the entire network of ground stations into communication, transmit instructions and control signals, and support service conversations during the operation of the satellite communications system.

12. The quality control system of the apparatus and communications channels which, in turn, consists of the complex of equipment for operative testing of the communications apparatus (transmitters, receivers, etc.) and a measurement complex, designed for tuning, preventive maintenance and testing of the quality of operation of the communications channels.

13. Various supplementary systems and nonelectronic services, such as the time service, power supply system, water supply system, air conditioning system, repair system, etc.

It is not difficult to imagine that the creation, organization and operation of such a complex system is an extremely

difficult task, requiring the joint efforts of specialists in various areas of science and technology. A particularly vital role is played by proper and well-founded selection of design, systems and organizational decisions, a result of technical planning, allowing the creation of highly effective satellite communications systems.

The primary trends, aspects and stages of this planning are presented in subsequent chapters of this monograph.

REFERENCES

1.1. Report in Technology Week, 1967, 19/IV, Vol. 20, No. 25, p. 30.

1.2. Hickman, V. D., "Canadian Plans for Creation of a National Satellite Communications System," Elektronika, 1967, No. 18, p. 57.

1.3. Journal des Telecommunication, July 1967, Vol. 34, No. 7, p. 234.

1.4. "Communique from the Conference of Experts of Socialist Countries, 5-13 April 1967," Moscow, Pravda, No. 106 (17,788), 16 April 1967.

1.5. Sistemy svyazi s ispol'zovaniyem iskusstvennykh sputnikov Zemli. [Communications Systems Using Artificial Earth Satellite], translated from the English, edited by V. V. Markov, Mir Press, 1964.

1.6. Crassner, G. N., J. V. Michaels, Vvedeniye v sistemy kosmicheskoy svyazi. [Introduction to Space Communications Systems], translated from the English, edited by M. G. Kroshkin and V. V. Markov, Svyaz' Press, 1967.

Chapter 2. Design of Satellite Communications Networks /21

§ 2.1. Operation of a Satellite Communications Network with Unassigned Channels

A satellite relay station, as we will show below, can support the operation of N_d duplex channels. The simplest method of utilization of these channels is to assign them to the S stations of the network. If we assume that all S stations should be connected with each other, i.e., communications are organized according to the "each with each" principle, any station should have S - 1 simplex communications links (each link may have several channels) for transmission of information to its correspondents and an equal number of simplex links for reception of response information. In all, the network should include l_s simplex channels or $l_d = l_s/2$. duplex channels. The number l_s can be easily determined, since each of the S stations transmits information over S - 1 lines; the entire network will include $l_s = S(S - 1)$ simplex links or $l_d = S(S - 1)/2$ duplex links.

By assigning the links and channels to definite stations, we produce an ordinary network, similar to networks constructed using terrestrial systems -- cables or radio relay links. In these networks, the channels are generally used comparatively poorly. For example, according to recommendation CCITT E51, the paid time of operation of one telephone channel in an automatic communications network averages 133.3 minutes per day or 34,000 minutes (6.5%) per year. The channels are used approximately the same amount of time for conversations between servicing personnel, and the remaining time they are idle due to the lack of load.

In communications networks using the delay-basis system, the channels are used better, but even in these networks the usage factor of assigned channels is comparatively low. Furthermore, one should keep in mind that the development of communications technology is moving in the direction of fully automated networks /22 and in the future the number of networks using the delay-basis system will continually decrease.

Thus, after creating a satellite communications system similar to ground systems, we will find that the load of the channels is extremely low, i.e., the system is uneconomical. At the same time, the cost of satellite communications systems is quite high at present, since the launching of satellites is difficult and

expensive, and repair of relay equipment which fails is at present simply impossible; the cost of ground stations is also quite high, particularly due to the large and complex antennas required, with their guidance and tracking systems.

This indicates that the problem of increasing the usage factor of channels is of decisive significance for increasing the economic effectiveness of satellite networks. A satellite network, since it covers a large area, allows a significant increase in the loading of channels in comparison to surface communications lines if unassigned lines and communications channels are used in organizing the operation of the network. Unassigned channels are those which can be used in the network by several stations in sequence: according to a schedule or as requirements for communications appear.

In one of the documents of the Eleventh Research Commission of CCITT (No. 73, 11 October 1966), the following terms are recommended, defining satellite communications channels according to the method of their use in the network:

1. "Pre-assigned communications channels among channels with multistation access," i.e., channels in the satellite communications network covering a certain territory and including several (more than two) stations. In spite of the word "assigned" in the definition, the channels are not permanently assigned, like land lines, but are rather assigned for a certain period of time. These channels are subdivided into two types:

Fixed channels, assigned in advance. Channels of this type are used in those cases when changes in requests during the working period are not significant and semiconstant assignment is desirable.

Channels assigned for definite times. These channels are used when considerable changes in the input flow of requests for communications are expected and channels must be assigned in a manner which changes with time; essentially, these are channels operating according to a schedule.

2. "Communications channels through satellites with multistation access, assigned on request."

This is a category of channels which can be used in those cases when the load in the network changes in a random manner.

Depending on the method of organization of work in the network, there may be three types of such channels:

-- "Fully variable channels," i.e., channels which can be organized between any pair of subscribers to the communications network; /23

-- "Variable receiving station channels." These are channels, the outgoing (transmitting) end of which is assigned for extended periods of time to some given station, while the incoming (receiving) end can be addressed to any subscriber of the network;

-- "Variable outgoing station channels," i.e., channels, the incoming (receiving) end of which are assigned to a given station for a certain time, while the outgoing end can be utilized by any station of the network as required.

Channels of the latter two types are also combined under the term "channels with multistation access with variable address."

Thus, there is today a rather broad classification of unassigned satellite channels. This is explained by the fact that these channels allow communications networks to be constructed which are different in principle from surface lines and are more economical.

To show the advantages of this type of operation, let us study several versions of the construction of a communications network with nine stations, connected according to the "each with each" principle.

As an example, we shall use the traffic of nine countries of Europe from the "general plan for development of the international network in the European zone and Mediterranean basin," prepared by the Combined Planning Committee of the CCITT and ECIR in Paris in 1966, "The Paris Plan of 1966." From this plan, we have taken the new, future streams, which must be supported between the nine countries by 1975, considering that half of these streams will be supported by surface communications lines, the other half -- by a satellite communications system.

To the figures for streams of telephone messages in the Paris Plan we have added 10% for telegraphy. Table 2.1 shows this traffic, with the volume of communications (in the numerator) indicated as the outgoing stream in minutes of communication per day. The same table, in the denominator, shows the number of channels which must be organized between stations if fully assigned channels are used as is done by surface networks. The number of channels was determined considering the fact that the load is distributed unevenly during the day and has at least one peak in the hour of maximum load (HML). The value of this peak is determined by the concentration factor, i.e., the portion of the total daily load accounted for by the HML. In Table 2.1, the concentration factor $\kappa = 0.15$.

Using coefficient κ, the flow at the HML is determined, then this flow is used to find the necessary number of channels, which can carry this flow with the required service quality [2.1], i.e., /24

with the required probability of failure p[1]. It is assumed in the calculation that p = 0.01. The total number of channels which must be assigned to each station and to the entire network as a whole is presented in Table 2.2 in column 3 and in Table 2.3 in column 2. As we can see, the network and, consequently, the satellite, must have some 400 duplex channels. It should be recalled that multichannel communications networks always utilize standard multiplexing equipment. In this case, obviously, communications between stations must be supported using twelve-channel communications groups, leading to a still greater number of required channels in the network (column 3, Table 2.3). True, the quality of servicing in the network will improve (probability of failure is reduced), but this improvement in quality above the planned level is not required.

TABLE 2.1

Country	France	West Germany	Yugoslavia	Bulgaria	Hungary	Rumania	Poland	Czechoslovakia	USSR
France	—	750 6	600 5	175 2	365 3	350 3	700 6	285 3	725 6
West Germany	750 6	—	465 4	470 4	1200 8	300 3	1650 10	3500 16	1250 8
Yugoslavia	600 5	465 4	—	825 7	450 4	200 2	180 2	385 3	525 4
Bulgaria	175 2	470 4	825 7	—	220 2	550 4	110 1	300 3	2250 12
Hungary	365 3	1200 8	450 4	220 2	—	525 4	385 3	675 6	1500 9
Rumania	350 3	300 3	200 2	550 4	525 4	—	175 2	300 3	1500 9
Poland	700 6	1650 10	180 2	110 1	385 3	175 2	—	600 5	2500 13
Czechoslovakia	285 3	3500 16	385 3	300 3	675 6	300 3	600 5	—	2000 11
USSR	725 6	1250 8	525 4	2250 12	1500 9	1500 9	2500 13	2000 11	—

[1]It is assumed that the communications network operates as a system with losses of calls.

TABLE 2.2

/25

Country	Total of Assigned Channels (Outgoing)	Total Outgoing Load, erlangs	Total Outgoing Load During HML, erlangs	Required Number of Channels in Trunk (Outgoing)
1	2	3	4	5
France	34	66	9.8	18
West Germany	59	160	24	34
Yugoslavia	31	60	9	17
Bulgaria	35	82	12.3	21
Hungary	39	89	13.3	22
Rumania	30	65	9.7	17
Poland	42	105	15.4	24
Czechoslovakia	50	134	20	30
USSR	72	205	30.7	42
Total:	392	966	124.2	225

TABLE 2.3

Indices	Numbers of Channels Necessary with Fully Assigned Channels	Same, Transmission in Standard Groups	Number of Channels Necessary in System and Channels with Variable Station Assignments	Same, Transmission of Standard Groups	Number of Channels Necessary Operating with "Completely Unassigned Channels"
1	2	3	4	5	6
Number of Channels in System	392	47 + 12 = = 564	225	9×60=540 or 22×12=264	140
Number of Channels in Relay Station	392	564	225	540 or 264	140
K	0.36	0.25	0.62	0.63	1

However, in a satellite communications network, the stations can be organized by other methods as well. Let us use the method of operation with "channels with variable station assignment." To do this, we combine the streams for each station into bundles which are transmitted by common group trunks. This method will be studied below in greater detail; suffice it to say at this point that the system can operate. Let us combine the streams for each station, thus increasing the channel utilization factor in the line due to the increase in this stream, making the system more effective. The results of calculation of such a system are presented in columns 3, 4 and 5 of Table 2.2 and columns 4 and 5 of Table 2.3, which show the total number of channels in the network and per satellite with this version for cases of use of nonstandard (column 4) and standard (column 5) group lines. We can see from the Table that the number of channels in the network decreases significantly in comparison with the case of fully assigned channels. /26

But the effectiveness of a satellite communications network can be still further increased by using completely unassigned channels. In this case, the entire volume of information to be transmitted through the network can be combined into a single stream, utilizing one powerful group of channels per satellite for its transmission[1]. In Table 2.2, the fourth column, bottom, gives the values of the total stream in this network: it is 124 erlangs. In order to transmit this stream of data with not over 1% loss of calls, a group of only 140 channels is required (Table 2.3). As we can see, the use of completely unassigned channels minimizes necessary network capacity for provision of any fixed servicing quality.

In order to describe the effectiveness of operation of the communications network, let us introduce the channel utilization factor K, which we will define as the ratio of the minimum necessary number of channels in the network $N_{\Sigma\ min}$, providing a fixed servicing quality (as is indicated above, it is attained with completely unassigned channels) to the number of channels N_{Σ} which must be available in the network with the method of operation used:

$$K = N_{\Sigma\ min}/N_{\Sigma}.$$

The bottom row of Table 2.3 presents the values of K for all cases in question.

Thus, in a satellite communications network, the use of unassigned channels can permit combination of streams into powerful

[1]This method of organization of operations in the network will also be studied in detail below.

bundles. This leads to a significant increase in the loading of the channels and, consequently, a decrease in the number of channels required in the network to provide a fixed servicing quality.

In our analysis, we have assumed that all stations in the network operate simultaneously, i.e., that the HML occurs in all countries at the same hour according to Central European Time. However, in actual systems this is not always so. If, for example, we analyze the communications network between Europe and America and between Europe and Asia, we find that the HML between Asia and Europe is quite different from the HML between Europe and America. In practice, in the morning, European time, communications will be heavy between Europe and Asia, and in the evening -- between Europe and America. In a network with assigned channels, it would /27
be necessary to have two communications links: Europe-Asia and Europe-America. In a network with unassigned channels, one link is sufficient, switched in the morning to Asia, in the evening to America (the capacity of the link should be sufficient, of course, to service either direction). Therefore, by using the time difference of the HML between different stations in the network, still heavier loading of the communications channels can be achieved, i.e., still greater effectiveness of network operations. The calculation of the gain which can be produced in this manner depends on the placement of the points of the network and their interactions, i.e., the traffic on the network, distributed through the hours of the day. This calculation can be performed easily in each specific case; the throughput capcity of the network and of a satellite can be increased by several times.

§ 2.2. Methods of Utilization of a Relay Station in the Design of Satellite Communications Networks

A communications satellite is an important element of the network, going a long way towards determining its structure. In organizing a network, various versions of utilization of the satellite relay station are possible (regardless of whether it is active or passive); selection of a method is determined by the requirements which must be satisfied by the planned communications network. The simplest version of use of a satellite in a network consists in directly connecting the communications stations to each other by independent communications links passing through the satellite (Figure 2.1). Each such link should be a duplex link, i.e., should contain two signals radiated by the two stations. The number of simplex channels r_{ij} in the link from station i to station j may differ, and is determined by the requirement of the station.

However, a satellite communications network with direct connection of stations over individual lines is not optimal. Its primary shortcoming is the necessity of providing a large number of

comparatively low-capacity independent communications links. The loss resulting from this structure of the network is double: /28

-- first of all, the poorer utilization of channels in the lines (small bundles of channels always have lower utilization factors);

-- secondly, the poorer utilization (for the same reason) of the power and bandwidth of the relay station.

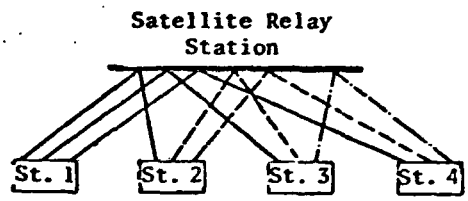


Figure 2.1. Direct Communications between Stations by Satellite Using Individual Communications Lines

Satellite communications systems allow the construction of significantly more effective networks. One version of such a network is a network with direct connection of stations utilizing not individual lines but powerful combined bundles (groups) of channels.

The operating principle of such a network is as follows: one of the stations of the network, for example A, radiates all its information in a common powerful bundle of channels (which we will call a multichannel subtrunk) to the satellite (Figure 2.2). This subtrunk simultaneously carries messages from all subscribers of this station, i.e., in the maximum version this subtrunk contains signals addressed to the (S - 1) other stations of the network. The signals are relayed by the satellite and radiated to the Earth, where they are received by these (S - 1) stations of the network. Each station (for example B), receiving a multichannel subtrunk, first separates it into individual

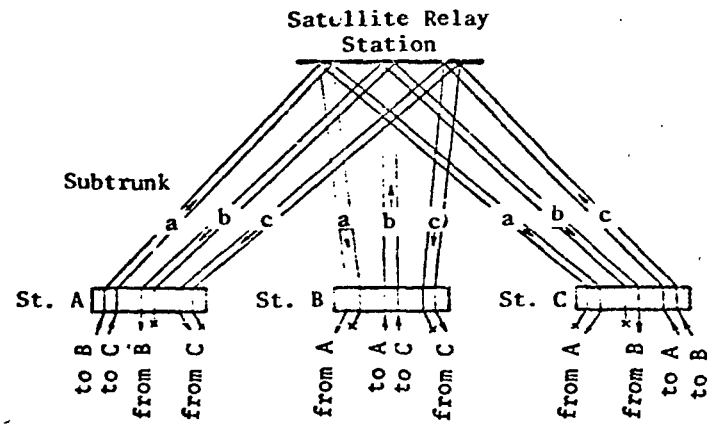


Figure 2.2. Use of Satellite Relay Station with Direct Communications Using Combined Bundles

channels, then, using a special control and switching system, separates out only those channels addressed to that station. In turn, answering station A, i.e., organizing a response channel, this station places it in the common bundle together with channels addressed to the other subscribers. With this method of operation, the network and satellite need have only S powerful multichannel simplex communications subtrunks, which is an advantage of this method. However, in this case each ground station must be able to /29 receive (before separation into individual channels) (S - 1) subtrunks, while transmitting only one.

Thus, the primary shortcoming of the method of direct communications using powerful combined bundles of channels is the poor utilization of the receiver apparatus of the ground stations.

Let us study yet another possible method of operation of a satellite communications system, which we will call communications through a center. In this case, it is also assumed that each station of the network, except for the one central station, radiates all the information intended for the other subscribers in a single common multichannel subtrunk (Figure 2.3).

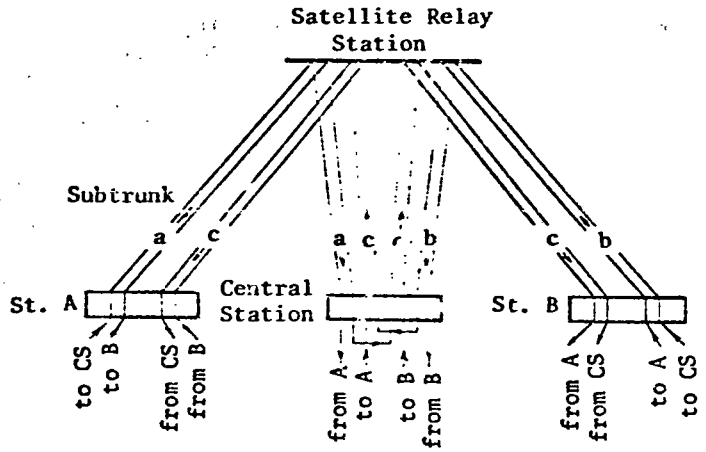


Figure 2.3. Use of Satellite Relay Station, Network Organized with Communications Through a Center

However, the signals of these subtrunks are received only at the one central station (CS). Thus, the CS receives information over the (S - 1) subtrunks from all peripheral stations (PS). The central station then performs complete separation of all subtrunks received into individual channels and then reswitches them. During the process of switching, groups of channels are formed, addressed to the various PS, which are then radiated from the CS in the form of multichannel subtrunks to the corresponding peripheral stations.

With this organization of the network, all peripheral stations must have reception and transmission equipment precisely corresponding to the stream of information to and from the stations:

$$Q_{tr\ i} = Q_{r\ i} = Q_i.$$

All the losses and complications which arise with this method of organization are shifted to the CS and the satellite relays station. A network utilizing communications through a center has two basic shortcomings. First of all, poor utilization of the satellite relay station; secondly, presence of two steps in the channels between the peripheral stations. /30

The low effectiveness of utilization of the relay station is a great shortcoming of this method of organization of a network, since the relay satellite is the most complex and expensive element of the system, the effectiveness of the utilization of which determines the economic effectiveness of the entire network to a great extent.

The presence of two steps in each channel becomes significant in those cases when the communications satellites move in high orbits, for example stationary orbits. A stationary orbit has an altitude of about 36,000 km above the Earth, and with two steps, the signal must travel a distance on the order of 170,000 km between one station and another, requiring almost 0.6 sec. With this great delay in signal propagation, duplex telephone communications becomes impossible. Since the telephony makes up the most significant fraction of any communications network, it is expedient to construct networks operating through the center only when low-flying satellites at altitudes of not over 10-15,000 km are used in the system. In this case, the delay in signal propagation is not too great.

Both of the methods of utilization of satellites in networks with multichannel communications lines studied above have shortcomings -- direct communications makes poor use of receiver apparatus in ground stations, while communications through a center makes poor use of the relay station on board the satellite. Combining these methods of constructing a satellite communications network, we can suggest a method of organization which is essentially free of these shortcomings. The combined version of organization is essentially a version with direct communication of stations in which, however, communications of peripheral stations among themselves and communications of these stations with the central station are separated.

Communications of PS among themselves is achieved by the version of direct communications. Each peripheral station radiates

information to all remaining PS in a single subtrunk (bundle of channels) and receives all subtrunks radiated by these PS, selecting from them those addressed to the given station. Communications of PS with CS is performed as in the system with communications through the center. For this, a certain number of subtrunks is assigned to the CS for each peripheral station, i.e., communications between PS and CS is performed so that each PS knows which of the CS subtrunks contains its channels and receives this subtrunk. Consequently, the PS do not receive all the channels of the CS as in the pure version of direct communications, but rather only one (or several if a standard subtrunk cannot transmit all of the information addressed to a given PS). The savings in receiving equipment of the PS is achieved due to the fact that it does not receive the subtrunks addressed to other PS from the CS. This savings will be greater, the greater the total flow from the CS to all PS, i.e., the more the planned communications network is centralized. Transmission from PS to CS is through a common subtrunk radiated from the PS. /31

Thus, we have studied the primary versions of organization of satellite networks from the standpoint of the method of utilization of the relay station. Of course, we can suggest other versions as well; however, all of them, like the combined method which we have studied, can be reduced to one of the methods studied or a combination of them.

However, still another method is possible, which differs in principle in its utilization of the satellite in the organization of a communications network. This is the organization of a network with multichannel communications lines using a satellite with switching equipment on board. In this network, each station radiates all its information to the satellite in one multichannel subtrunk. On board the satellite, the information from all stations is received, divided into individual channels, which are switched according to the addresses of the messages they carry, and new multichannel subtrunks are formed, directed from the satellite, each to its corresponding station. In this network, both the satellite and the ground equipment would be used completely and from this standpoint the system would be optimal. However, the satellite relay would be converted to such a complex, cumbersome and expensive structure that it would hardly be possible to create such satellites in the forseeable future.

Comparing the versions studied, we must keep in mind that each has a number of advantages and shortcomings, defining their area of use in planned satellite communications networks.

§ 2.3. Provision of Broadcast-Type Satellite Communications Networks

Of the entire variety of satellite networks, we will describe in this section some of the main versions in order to show the capabilities of such networks, and also to allow us to study methods of design of satellite communications networks using these versions.

First of all, let us turn to so-called broadcast networks. By a broadcast communications network, we refer to a network (Figure 2.4) in which at any given moment in time, one station is transmitting (radiating) information, while many stations receive the information. No stations receiving the information can answer, i.e., there are only simplex (one-way) communications links in the network. The best example of such a network is a network used to transmit television and radio broadcasts. /32

A broadcast network always has a central station which controls its operation. Most frequently, the broadcast information is transmitted from the central station of all the remaining points in the network. However, there may also be several transmitting centers in the network, which operate in turn, or even sometimes simultaneously, for example if television programs are exchanged between two countries included in a common television broadcast network. In this case, each receiving station, depending on its capabilities and desires, may receive one of the programs transmitted or both. The organization of the operation of such a network is quite simple. The center (or centers) transmit information, while all receiving stations of the network receive it. Sometimes it is possible for the center to operate not with all stations simultaneously, but with individual groups of stations. It is also possible for several transmitting centers to operate with groups of stations, for example, in the transmission of television to various countries included in a common network but having their own transmitting centers, in addition to one common center for the entire network.

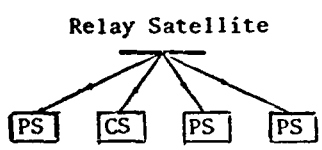


Figure 2.4. Broadcast Satellite Communications Network

Sometimes the center can transmit one program in sequence to individual groups of stations. For example, when television programs are transmitted over a large area, one program may be transmitted at first to a group of the easternmost receiving stations, where the local time is most convenient for reception of the program; then, two or three

hours later, the same program might be repeated for stations further west and, possibly, two or three hours later it might be repeated once more for the westernmost stations of the network.

The operation of the channels in a broadcast communications network is controlled by the central station either according to a schedule or as requirements for information transmission appear. The network must have a service communications system, over which the CS can transmit its instructions to all PS. The service communications system should, of course, be duplex. In this case, the network may either have auxiliary duplex satellite communication channels or auxiliary channels can be organized over surface communications lines.

As a rule, purely broadcast satellite networks are not planned; therefore, the possibility is always provided of arranging service communications through the satellite, particularly since the service communications required for a broadcast network are very small in volume. Basically, they consist of instructions to connect and disconnect equipment and certain supplementary messages concerning the quality of operation, changes in schedule, etc. /33

Thus, provision of broadcast satellite communications networks is simple, which is a great advantage. In order to transmit the same circular information over surface circuits, as a rule, much more complex organization is required, since it is always necessary for the information to pass through a number of switching centers, to interact with other types of information transmitted through the same circuits, etc.

§ 2.4. Design of Low-Channel Centralized Networks

We will study the arrangements for "low-channel" and "multichannel" satellite communications networks. The term low-channel network does not refer to the throughput capacity of the network as a whole or of its individual stations, but rather to the capacity of independent communications subtrunks.

In low-channel networks, we will always assume that each subtrunk has a capacity on the order of one to three channels, this capacity most frequently being one channel. There may be a large number of such low-channel (or single channel) subtrunks in the communications network -- up to several thousand; each station of the network may cover dozens or even hundreds of subtrunks but, from the point of view of organization of operations, these networks have one characteristic peculiarity, distinguishing them from other satellite communications networks, in that each subtrunk (single channel or two-three channel) has an inseparable, independent communications link, which is controlled and switched in its entirety and used by any pair of subscribers.

The term "multichannel network" also refers not to the capacity of the network or of its stations, but rather to the capacity of its subtrunks, which the network uses in operation. Multichannel subtrunks have capacities of from 12 channels upward and are formed using multiplexing equipment. Furthermore, as will be demonstrated below, multichannel subtrunks in satellite networks are no longer inseparable, as in surface lines, but rather each such subtrunk can be utilized for communications by several stations simultaneously.

Let us now turn to the study of low-channel communications networks. The organization of low-channel communications with centralized control is simplest. In this type of network, there are l_s simplex, completely variable (unassigned) low-channel communications channel Since, generally, information exchange is duplex, /34
the communications lines are paired in advance, so that we can state that the network has l_d duplex communications lines of $r = 1\text{-}3$ channels each. The communications lines are at the disposal of the network control center (CC).

Each station of the network has transmitting-receiving and terminal apparatus, designed to operate over any possible line. For example, if the links differ as to carrier frequency, each station must have receivers and transmitters capable of being tuned to the frequency of all the communications links in the network; if the communications system utilizes time-division multiplexing, the transmitting and receiving equipment must be capable of operating in each of the time intervals set aside for individual lines. The amount of equipment at each station depends on the flow of information which the station is to transmit.

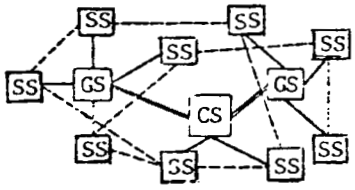


Figure 2.5. Low-Channel Satellite Communications Network

A low-channel communications network (Figure 2.5) generally has a central station (CS) and a large number of peripheral stations. The traffic at the central station is heavy, since it is connected to all peripheral stations; the traffic of most peripheral stations, on the other hand, is limited. A peripheral station is most frequently a terminal station and requires one to two lines. These peripheral stations will be called subscriber stations (SS). In addition to CS and SS, the network may also include stations connected with a given group of subscribers in an area served by this station and having rather large traffic volume. These

stations, which are communications junctions, will be called group stations (GS).

This network is externally somewhat similar to a radial-junction network of surface communications lines, but differs from it in principle in the method of organization of channel operation. In the satellite communications network, all stations -- CS, SS and GS -- can be connected with each other according to the "each with each" principle, and directly, whereas in surface networks, these connections pass through a number of switching centers. The presence of switching centers is characteristic for surface networks, since they are used to form the large bundles of channels, which are highly effective. For satellite low-channel networks, these switching centers are not required, since the entire network has only one common large bundle -- the bundle of channels through the satellite, which communicates with all stations in the network and is therefore particularly effectively used. The presence of the central and group stations in the satellite network results from economic, geographic or administrative factors, not from the requirement for organizational construction of the network. /35

Setting up of communications channels in a low-channel network results from the requirements of a subscriber. When a request appears at station A for communications with station B, this station sends a request over the service line to the CC (Figure 2.6). At the control center (CC), there is a memory system which carries the distribution (status) of all communications lines of the network, as well as the status and capacity of all stations. Upon receipt of a request for a channel from station A to station B, the CC analyzes the situation -- seeks a free line -- determines whether station B can obtain a channel with station A (whether there is equipment free at the station for organization of the channel). If it is found that it is possible to obtain such a channel, the CC sends out an instruction to stations A and B to switch equipment to the proper communications circuit (at a given set of frequencies, time intervals, etc.). The equipment should be automatically switched. If the CC determines that for any reason it is impossible to set up a communications line as requested, for example there are no available communications lines or the subscriber called is busy, the busy signal is sent to station A.

After setting up of a communications circuit, it is noted in the /36
memory of the CC that the line is busy, and the time of beginning of operation is also noted. After the conversation is completed, the subscribers send the completion signal through the service lines to the CC. The CC then notes that the line is free and the time of completion of the call, which can be used to bill the subscribers for communications time. The information that the line is free is recorded in CC memory, and the line can then be used by any other subscribers. We can see from the algorithm of network operation that the service channels between the CC and all stations

of the network are quite significant. These channels should operate continually in the duty mode and independently of the operation of the main communications lines.

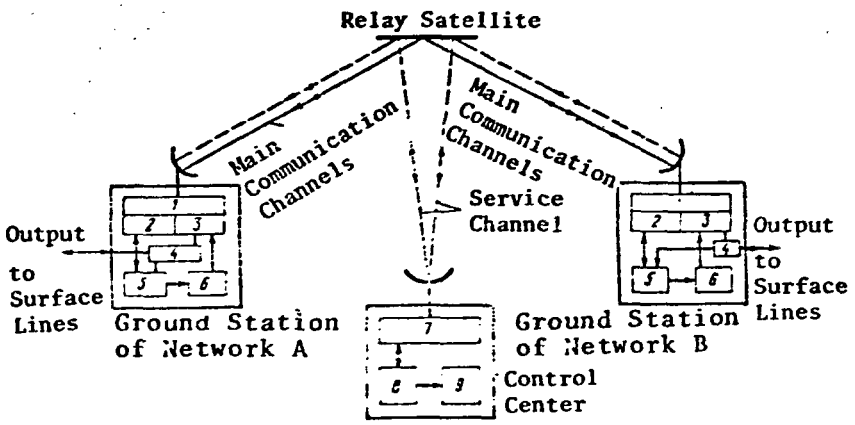


Figure 2.6. Communications in a Low-Channel Communications Network: 1, Transceiver; 2, Service Channel; 3, Main Channels; 4, Terminal Apparatus; 5, Connecting Equipment; 6, Channel Selecting Equipment; 7, Service Communications Transceiver; 8, Control (Logic) Devices for Connection; 9, Memory Device Carrying Situation in Network

A low-channel satellite communications network with centralized control has been planned as a part of the STAR American-Japanese system [2.2-2.6].*

This network utilizes single-channel communications lines. All lines pass through a common satellite relay station and differ in carrier frequencies, i.e., multistation access is used with frequency division of subtrunks. In order to facilitate the organization of channels, the communications lines are paired into duplex channels in advance; each duplex channel consists of two lines at two carrier frequencies. The network includes a control center which distributes the channels upon subscribers requests. Each station in the network has continuous service communications with the CC.

The number of stations in the network and their throughput capacity are not limited, although it is assumed that there will not be over 100 stations. The network, according to the planners, is to cover the entire Earth, i.e., will be global. It is to

*Translator's Note: Reference 2.3-2.6 are not given in the original foreign text.

include many countries of the world, each of which must construct one or more ground stations for this purpose. The capacity of the stations will be determined by the traffic and may be quite great -- up to 100 channels.

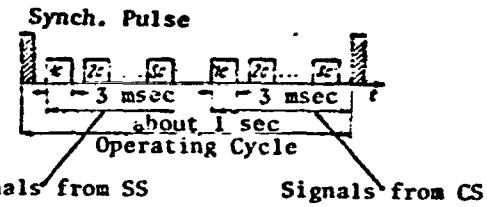


Figure 2.7. Operating Cycle of Service Communications Line in the STAR System

The operating algorithm of the network, i.e., the procedure for occupying a channel, agrees fully with that described above (Figure 2.6).

The organization of service communications in the STAR system is interesting. The system utilizes one special data transmission channel for this purpose, in which information is transmitted in binary form at a rate of 25,000 bps. The service channel can be used in sequence by all stations of the network and the control center. For this purpose, the service channel operates in cycles. The duration of one cycle is approximately one second. At the beginning of the cycle, the CC transmits a synchronizing pulse, used by each station to calculate its moment for communications, i.e., the moment when it can utilize the service channel for transmission of its messages to the CC (Figure 2.7). During the course of a cycle, each station in the network can utilize the service channel once for transmission of a request for one communications channel. Each request is transmitted as a sector of binary information containing approximately 80 bits and lasting about 2-3 msec. The CC can utilize the service channel for transmission of answers to the stations during equal time intervals. During the course of a cycle of operation, the CC utilizes the service channel as many times as requests have been received.

It is calculated that with this organization of service communications, a station with up to 30 communications channels can support their operation with the required service quality. If a ground station has over 30 communications channels, it will be assigned more than one time interval during one cycle of operation. /37 Three intervals per cycle are sufficient to serve 200 communications channels.

The transmission of telegraph messages can be performed over telephone communications channels using secondary multiplexing, as is done in surface communications lines. However, this utilization of telephone channels is not desirable with this type of organization. In this case, the band of one simplex communications line -- about 100 KHz -- can be used for the organization of only 24 telegraph channels.

In the STAR system, telegraphy is to be transmitted using special high-speed data transmission lines operating at 25,000 bps, similar to the service communications channel. These channels, like the service channel, can be utilized by all stations of the network for predetermined time intervals.

However, in this case the channel is occupied not in sequence but in time intervals strictly assigned in advance. The telegraph channels are assigned by request, as are the telephone channels. The interval during which a given station can use the telegraph channel is assigned on request by the central control station.

Telegraph messages are transmitted at high speed (25,000 bps), requiring that the stations store a significant volume of messages in advance. With this type of operation, one data transmission channel, occupying a bandwidth of less than 100 KHz in the satellite, can support the operation of about 500 telegraph channels.

Great attention has been turned in the development of the STAR system to its interaction with the networks of surface communications lines. The satellite communications lines form a part of the channels connecting to subscribers. In addition to the satellite network, these channels pass through ground network switching stations, which operate according to certain algorithms and utilize standard, generally accepted interaction and signaling symbols in their organization. The satellite communications network requires a number of new operations for the organization of channels, requiring new interaction signals. In order to avoid disrupting the operating order of the international communications networks already established, the system uses special transition apparatus at all ground stations, receiving signals in the generally accepted signaling system from the surface communications networks and converting them to signals required for functioning of the satellite network. In just the same way, the interaction signals used within the network are reconverted to the standard surface system signals as the communications lines leave the satellite network.

A satellite communications network with centralized control can be constructed by another method as well. In the version we have just analyzed, a station occupies a channel after requesting it from the CC, where the status of the network is first analyzed, then a decision is made concerning organization of a channel. With this working algorithm, the entire process of organization of a channel occurs after the request for a channel arrives at the CC.

However, the CC can prepare free channels for communications in advance, before requests appear.

In this case, the free channels are marked in some manner, so that any station needing a channel can determine whether the network has free channels available without turning to the CC, and

take the channels independently, making the connection with its required subscriber. Thus, the role of the CC in this version is reduced to preparation of free channels and to helping the stations when these channels are busy. No direct conversations occur between the stations and control center in this type of network, and the stations make their own connections independently. The network does not require service channels between the CC and the stations. The entire procedure of occupying a channel can be performed in this case over the main (conversation) channels. /38

Let us study the operation of such a network in greater detail. We will call it, in contrast to the network we have just described, a centralized network with direct establishment of connections between subscribers (Figure 2.8). The network and the satellite have a certain number of paired (duplex) unassigned communications lines 1_d. The control center checks the operation of these lines, determining which are free at each given moment. All free duplex lines are marked by the center. Marking consists in that a special marker signal is transmitted over one of the two paired lines by the CC, indicating that the duplex line is free. For definition, it is established that of the pair on the ith duplex line, consisting of two simplex lines with numbers i and i' (i = 1, 2, 3, ..., 1_d), the CC always marks the first lines, i.e., the ith line.

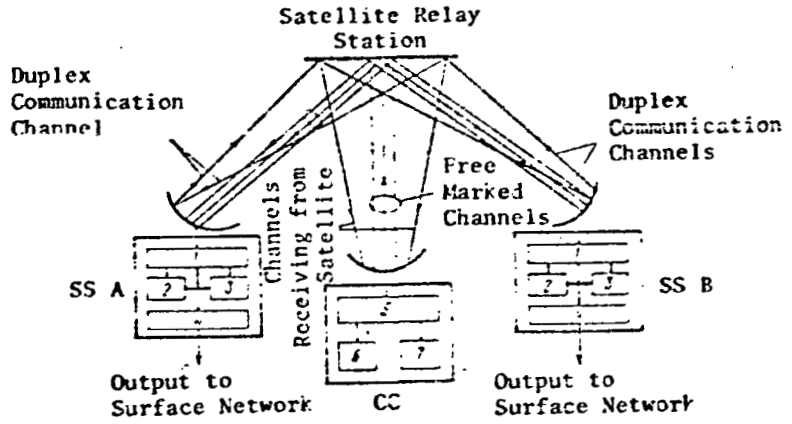


Figure 2.8. Centralized Network with Direct Establishment of Connection: 1, Transceiver Apparatus of SS; 2, Device for Search for Free Channels and Establishment of Connection at SS; 3, SS Channel Selection Apparatus; 4, SS Terminal Apparatus; 5, CC Transceiver Apparatus; 6, Device for Establishing Connections at CC; 7, CC Free Channel Marker Device

The receiver devices of all stations of the network operate in the duty reception mode, tuning in to each line of the network in turn. This requires that the receiver at the station be continually retuned to each of the simplex lines which may be marked (with numbers i). The receiver may be retuned either strictly in sequence, if the number of channels in the network is not great or, if the total capacity of the network is great and the servicing cycle of all channels would be too long, parallel-sequential tuning /39 can be used, inspecting several channels simultaneously. Of course, this requires that the station have several duty receivers.

If station A must be connected to station B, at station A the instruction is given for the duty receiver to find a free channel. On this instruction, the receiver stops at the first channel with number i over which a marker is being transmitted. Immediately after this, the transmitter of the station is turned on and switched to the partner line with number i'. The line busy signal is transmitted over this line to the CC, which then removes the marker from line i, indicating that it is busy. For greater reliability, the CC may also transmit a brief special signal confirming the assignment of the line.

After the receiver of station A receives confirmation of assignment of line i, the transmitter of this station is switched from line i' to line i and transmits a call for station B. The call signal must be sufficiently long that the duty receiver of station B will receive the call during its process of inspection of the channels of the network. The call signal from station A to station B contains the address (number) of station B, a signal representing a call and the return address (number) of station A, which station B uses to determine which subscriber desires communications. Upon receiving the call, station B stops its receiver at line i and turns on its transmitter, adjusted to the frequency of line i'. Then station B transmits a confirmation of the reception of the call signal of station A over line i', after which the line between A and B can be considered ready for operations. After completion of communications, station A, the initiator of the call, sends the completion signal to the CC, after which the CC once more marks line i by radiating the special marker signal.

If the call to station B is busy and cannot respond to the call of station A, in the simplest version it does not receive the call (the receiver does not receive call signals, since it is busy with other communications) or it receives the call (using a special duty receiver) but does not send the confirmation of call reception to station A. In any case, station A, when it does not receive confirmation from B that its call has been received, makes a repeated attempt to organize communications with B after a certain period of time has passed.

As we see, in a centralized network with direct establishment of connections, the functions of the control center are much

simpler than in a network with centralized control. This is a great advantage of this method of organization. If the CC fails in a network with centralized control, the network ceases operation, since it is the center which organizes channels (even including remote switching and tuning of equipment). Of course, the central station can be replaced with a reserve center, but the CC /40
is a rather complex and expensive device. Furthermore, in order to put a reserve CC in operation, it must know the entire situation on the network, know which connections have been made by the operating main CC. In fact, in order to assure failure free operation of the network in case of failure of a CC, two control centers must operate on the network continually -- the main and reserve centers -- but the reserve center, performing all operations involved in control of the network, does not transmit the instructions which it develops, but only coordinates them with the instructions developed by the main CC. If the main CC fails, the reserve CC easily begins operating the network in place of the main CC, by simply beginning to transmit its instructions into the network.

It should be noted also that in a centralized network with direct establishment of connections, the CC is much simpler: it does not require a cumbersome memory system to record the situation in the entire network or a complex system of logic devices, which must analyze the situation each time communications is established. The CC need only have a free line marker system and very simple logic system, which connects and disconnects the marker signal after the busy and completion signals are received from the stations. Replacement of this type of CC is a much simpler affair, particularly since the CC can comparatively easily analyze which lines in the network are free once more, so that parallel operation of the two CC is not required.

One shortcoming of this method of organization is the complexity of checking for free channels and reception of calls at stations, particularly if the network utilizes many channels, which must all be checked. Performance of this rapid testing of all channels at the stations requires cumbersome tracking devices. Furthermore, the addressing system will also be rather complex, since the address must be received accurately in a limited period of time, the time during which the duty receiver inspects each channel.

Another shortcoming of a centralized network with direct establishment of connection is the possibility of additional losses as communications are established, which may arise due to the fact that several stations attempt to occupy the same channel simultaneously. Actually, as we know the propagation time of signals in satellite lines is quite great, reaching 300 msec when stationary satellites are used. Consequently, if any station A wishes to occupy a free line, the marker will be removed from this line only 300 msec after the station sends out the busy

signal. However, another station, say B, will receive the information that the marker has been removed from the line only 300 msec after it is removed at the CC which, consequently, is 600 msec after the line is actually taken by station A. During these 600 msec, station B may attempt to occupy the line which is already busy, may interpret the removal of the marker and the confirmation sent by station A as a response to its own call and may begin communications over the same line as station A. As a result, neither station can operate normally with its desired correspondent and additional losses appear in the network. The magnitude of these losses will be calculated later, but let us now note simply that several methods to reduce them can be suggested. If the signal propagation time in the network is constant when, for example, a stationary satellite is used, a delay of 600 msec can be inserted between the moment of transmission of a busy signal to the CC and the moment of reception of the signal confirming that a given line is busy from the CC. A delay between these moments of less than or greater than 600 msec means that the CC has responded (by removing the marker) to the request of another station. In this case, the station should not occupy this line, but must seek further for another line. With this operating system, losses due to simultaneous occupation of a line by several stations are sharply reduced, although it is hardly possible to eliminate them completely. /41

This version of the operation of a centralized network with direct establishment of connections is not the only possible version. Various modifications of this version can be suggested. One is worthy of particular attention. In this case, the CC marks not all the available free lines, but rather only one. In this case the stations of the network need not continually inspect all lines of the network in the duty reception mode as was described above. With this method, all stations receive the same line, and the line is taken by the station which needs the line at the moment. As soon as the line is taken, the CC marks the next free line, and the receivers of all stations are switched to this line. In this mode, the operation of the duty receivers at the stations is significantly easier and the free line search time is reduced, although the probability is increased that the free line will be requested by several stations simultaneously, i.e., the probability of losses increases. The losses can be reduced to some extent if not one, but several (2-3) lines are marked simultaneously, and each station inspects these marked lines in sequence. Probably, this modification is more expedient than the original version if the number of free lines simultaneously marked is properly selected.

In speaking of low channel satellite communications networks up to this point we have assumed that all stations of the network are identical in their power (electronic) indicators and differ only in their throughput capacity, i.e., the number of lines used for communications. It has also been assumed that the signals

radiated by all stations of the networks of the satellite are identical. In those cases when the satellite communications network includes stations with comparatively greater streams of information, assigned to points with constant information exchange -- cities, large villages, etc. -- this situation may obtain. In areas with heavy traffic, it is sensible to construct stations with comparatively larger antennas, assuring good quality of operation of the channels and, as a rule, reception of radio broadcast and television transmissions. /42

However, it may be necessary to construct a satellite communications network servicing a large number of small points, some of which may even be mobile -- for example, a satellite communications network for large construction projects in remote areas, expeditions, or a network servicing merchant marine ships, etc. In these networks, it is expedient to organize the communications system differently. Keeping in mind that the traffic of the stations of the network will not be great, that the stations primarily transmit their information to the center, that the cost of stations, their complexity and dimensions should be as small as possible, it is expedient to construct a communications system such that the central and group stations (CS and GS) differ in their electronic parameters from subscriber stations (SS). CS and GS in these stations have large antennas and radiate powerful signals to the satellites, while SS, on the other hand, have small antennas (2-5 m in diameter). The powerful CS and GS stations can be received at subscriber stations by small antennas, while the weak SS signals will be successfully received at CS and GS, which have large antennas. In this network, the SS will be significantly less expensive than CS or GS and, therefore, the cost of a communications channel for them will be smaller.

However, in this sort of network direct communications between SS is impossible due to power considerations. In order for SS to be able to exchange information, they must talk with additional relay, i.e., in two steps. If the system utilizes high-altitude satellites, the quality of communications over these links will be poor. However, if few such conversations are required, it is sensible to construct networks using this principle, considering their significant economic advantages. Furthermore, SS which must maintain constant communications with each other can be made with large antennas similar to CS antennas, or with more powerful transmitters In this case, these particular SS can organize direct communications with each other.

A low-channel communications network with small SS is particularly convenient for construction according to the principle of centralized networks with direct establishment of connections. Since in these networks the SS must contact the CC to establish a connection, by placing the CC at a central (or group) station,

the process of making a connection can be significantly simplified and accelerated.

In this case, the CC, receiving the busy signal from an SS, /43 removes the marker and at the same time connects the SS to free channels at the CS. After removal of the marker, a signal indicating that a connection has been made and the line is ready for operation is immediately sent from the CC to the SS.

If two CC must be connected through the center, the CC, after receiving from SS-1 the busy signal and the number of SS-2, with which SS-1 requires communications, sends out a call to SS-2, establishes the connection with it, then switches the channels of SS-1 and SS-2 (Figure 2.9).

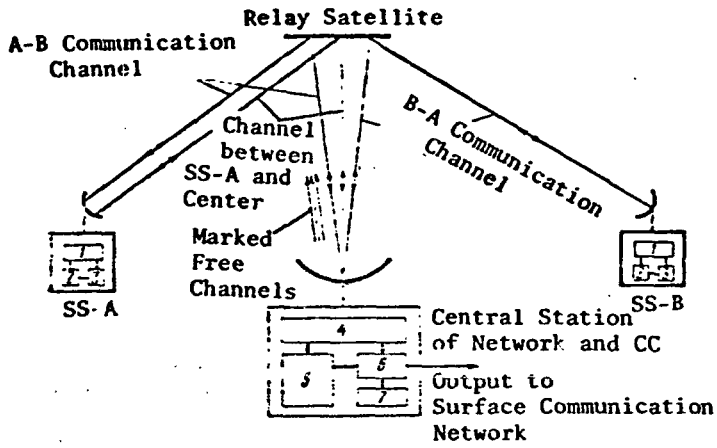


Figure 2.9. Low-Channel Communications Network with Small Subscriber Stations: 1, SS Transceiver Apparatus; 2, Apparatus for Establishment of Connection (Control) at SS; 3, SS Terminal Apparatus; 4, CS Transceiver Apparatus; 5, Apparatus Marking Free Channels and Establishing Connections at CS; 6, CS Terminal Apparatus; 7, CS Switching Apparatus

§ 2.5. Low-Channel Decentralized Networks

It was mentioned above that the presence of a single control center in a network has its disadvantages. First of all, with centralized control of the network, its operational reliability is reduced, since failure of the CC disrupts the operation of the entire network. Furthermore, the CC restricts the actions of the subscriber stations somewhat, reduces their independence and,

since it may be busy at times, introduces additional losses of calls. Finally, the control center is a rather complex and expensive structure, increasing the frequently very high cost of construction and operation of a communications network.

These factors indicate the necessity of constructing networks in which the control center is completely absent, i.e., networks with decentralized control. In these networks, establishment of a connection between any two stations is performed by the stations themselves; no one, except these stations, takes any part in the process of establishment of a connection. /44

Consequently, a low-channel communications network with decentralized control consists of a number of stations with equal rights, differing from each other only in the number of channels used. The stations should be identical in their electronic parameters, although in principle radial connections may predominate in a network with decentralized control. However, it is hardly expedient to construct a decentralized network with small and large stations operating with different signals (as was described above) and purely radial connections, since this leads to unjustified complication of SS.

Each station in a network with decentralized control should be able to organize channels with any other station of the network or, at least, with any station in some given group. When a network is organized with decentralized control, two main problems must be solved, which are characteristic for this method of operations. These include first of all the problem of control of the load on the satellite relay station.

In a network with centralized control, this problem does not arise, since the CC, no matter what the method of control, assigns the stations free channels if available on the satellite (and in the network). With decentralized control, the picture changes. Here the stations attempt to occupy communications lines independently, without knowing whether free lines are available on the satellite. The second problem of decentralized control, related to the first, is the problem of seeking out a free channel on the satellite and determining whether the station called is free, whether it can answer the call.

We will study the manner in which these problems are solved using the example of a decentralized network with channels assigned at one end. Suppose the communications system is constructed so that one-channel communications lines are used. A certain number of such lines are assigned to each station at the receiving end. The assignment consists in that the station is assigned several addresses by which it can be called. The address might be the frequency of signals received if lines with frequency division are used in the network, or a certain combination of frequency and time intervals if a communications system

with frequency-time division is used, etc. It is important that the address in the network define the communications channel. Selecting an address, a subscriber occupies a given communications channel which is assigned (at the receiving end) to some station.

Let us further assume that the number of signals which the satellite can relay is N_s, the number of addresses which can be placed in the communications network $N_a > N_s$. This situation is quite realistic and means that the power capabilities of the satellite are less than the capabilities of the addressing system in the network. /45

For definition, let us study a network in which frequency division of channels is used, i.e., a network in which each simplex channel utilizes a separate carrier frequency. The frequencies of carriers are assigned to stations for reception and, consequently, form the addresses of the stations. The condition $N_a > N_s$ means that the relay system can transmit not over N_s signals simultaneously, while its transmission pass band allows N_a frequencies, i.e., N_a addresses, to be used.

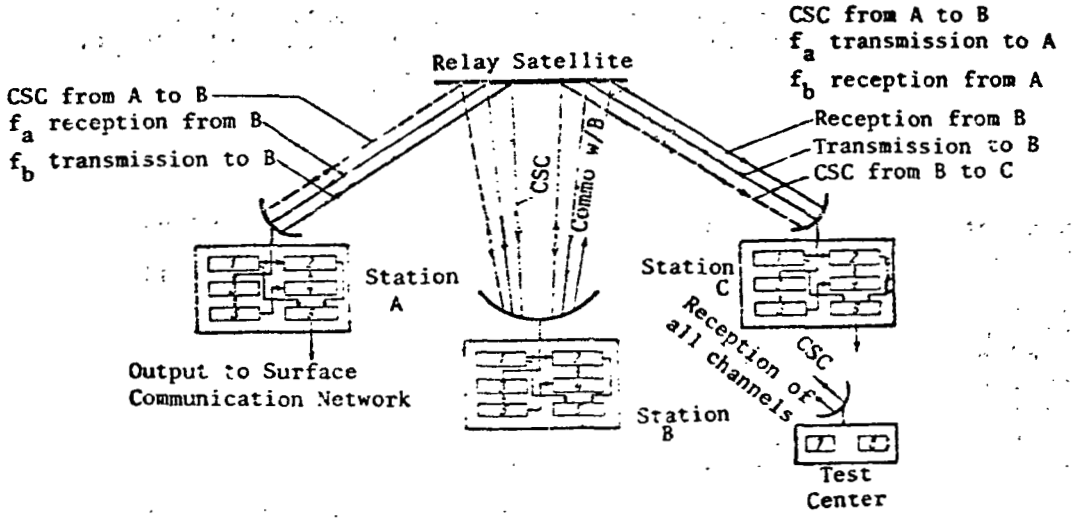


Figure 2.10. Low-Channel Network with Decentralized Control: 1, Receiver with Channels Assigned by Frequencies; 2, Transmitter with Channels Adjustable by Frequencies; 3, CSC Transceiver; 4, Transmitter Channel Frequency Selector; 5, Connection Device; 6, Terminal Apparatus; 7, TC Working Channel Receiver; 8, Transceiver for CSC with TC

If these conditions are fulfilled, a communications network can be organized with decentralized control (Figure 2.10). One or more frequencies is assigned for each station (from f_i to f_j) for reception; consequently, in order to communicate with this station, these frequencies must be transmitted to it. In order for all stations in the network to be capable of communicating with each other, the transmitters of the stations are made adjustable throughout the entire range. Thus, station A, radiating frequencies assigned to station B, can generate the required number of channels, since B can radiate the response signals on the frequencies of station A.

The number of signals which can be transmitted to the satellite /46
N_a is greater than the power capabilities of its relay station. If the stations of the network begin to exchange information in all channels simultaneously, the relay station will be overloaded and the network cannot operate. However, the probability of this situation in a properly planned network is quite slight. If we know the communications requirements of the stations, i.e., we know the mean daily load of each station in call-hours and the primary operating times, we can calculate the probability of an overload of the relay station. Therefore, in planning a network, based on the statistical nature of the process of information exchange, we can determine the number of stations in the network so as always to assure normal (without overload) operation of the relay station with any predetermined probability. Finally, it is desirable to have some check of the load of the relay station in order to allow the operation of the network to be altered in case of excessive load, to restore the normal mode. Since any network has a quality control system, it is at this test center that the load of the satellite is tested. Overloading of the relay station causes noise to appear in the channels at a higher power level than called for in the norms. After detecting this increased noise, the test system should reduce the number of channels passing through the satellite by sending an instruction to this effect to various stations. Incidentally, as will be shown below, with proper design of the network, overload of the relay station can be almost completely eliminated and it will hardly ever be necessary to reduce the information exchange artificially.

Thus, the first problem typical for a decentralized network -- the problem of overloading of the satellite -- is solved by proper selection of the traffic of the network. The second problem -- determination of free and busy channels -- is solved by efficient organization of the network. It is most convenient to introduce special call and service channels (CSC), using which the connections are made. The algorithm for establishment of a connection will then be as follows. First station A, calling station B, sends the call signal for station B through the CSC, reporting its return address in the signal. Station B, when it receives the call (for

which the stations always receive the CSC), sends station A either confirmation of reception of the call and the frequency of the channel on which it will answer station A, or the "busy" signal, depending on the situation. In order to decrease the load on the CSC, station B can send its answer to station A not over the CSC, but over the channel which station A has prepared for communications with station B. The "busy" signal, of course, must be sent through the CSC.

There may be either one CSC or several in a network; each CSC serves its own group of stations (channels). If one CSC serves several stations, these stations work in a service channel using /47
a single common frequency. They are differentiated by their addresses transmitted through the CSC before transmission of the call signal. Access to the calling and service channel by the stations which it serves should be free. Considering that messages transmitted through the CSC are very short, it is always possible to calculate the number of stations to be served by the CSC such that the probability of its simultaneous occupation by two or more calls is negligible. If, still, two stations try to use the CSC simultaneously, the calls are lost and they must repeat their attempt to establish connection.

It must be noted that in the network which we have described it is not necessary to assign the receiving ends of the channels, i.e., construct the system with "variable calling station." It is just as suitable to assign the transmitting ends of all information channels and have a system with "variable called station," in which case it is the receivers rather than the transmitters which must be tunable over a range. It is important only that the calling and service channels always be assigned as to reception, since in this case it is easy to assure continuous monitoring of the calling channels.

In planning a decentralized network operating according to this principle, it may be found that the number of frequencies at the relay station is less than the number of channels potentially necessary, i.e., that it is impossible to assign one frequency to each channel, since there are not enough frequencies. In this case (Figure 2.11), one frequency may be assigned to two, three or more channels, thus increasing the number of channels in the network. However, this expansion of the network is accompanied by deterioration in service quality. Actually, if each channel is assigned one frequency, the communications network is fully accessible -- any pair of subscribers can organize a channel for communications regardless of the status of all remaining lines and stations in the network.

If any frequency is assigned to two channels, i.e., two subscribers, one of the subscribers must wait if the other subscriber has taken their common frequency. Thus, we produce a network in which the organization of a channel between any pair of subscribers

depends on the status of the network, i.e., on whether other subscribers are talking. This type of network is incompletely accessible, and additional possibilities of loss of calls arise, which must be considered in planning.

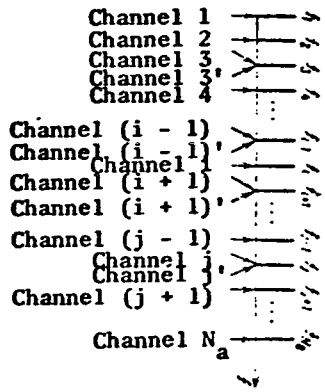


Figure 2.11. Distribution of Frequencies in Decentralized Network with Incomplete Station Access

Other versions of organization of decentralizated low-channel communications networks can be suggested, which do not have this shortcoming. For example, a network can be constructed similar to a centralized network with direct establishment of connection, but without marking of the free channels at the CC. /48

In this case, each station, in the process of search for a free channel, simply inspects all channels of the network and determines whether a given channel is free or not on the basis of the absence or presence of transmitted information (a signal) in the channel. This method of operation is quite possible in principle, but is quite inconvenient in practice, since the process of establishment of a connection will require a long time, and the probability of erroneous occupation of a channel already busy is rather high.

Probably, the version of a decentralized network described above must be considered one of the best possible, and in our subsequent presentation we will analyze only it.

We should emphasize once more that a network with decentralized control can use various methods of multistation operation: frequency division of channels, time division or shape division. Here and in the following we will study only the version with frequency division of channels in order to make our presentation more specific. The algorithm for control of the network and the conclusions produced above relate fully to all other versions as well.

§ 2.6. Organization of Multichannel Network with Communications Through a Center

Let us study multichannel satellite communications networks, i.e., networks in which the stations radiate and receive multichannel communications subtrunks (of 12, 60 or more channels). As

before, the network will operate with unassigned channels, switched on request. In the simplest version, the network can operate only with switching of undivided subtrunks, which are not assigned and are allocated by the center as required by the stations. In this case, we have an ordinary network with centralized control, similar to that described above, but lines in this network are multichannel lines. The operation of the network is similar to the operation of a low-channel network. However, the servicing quality and effectiveness of channel use in this network are much poorer than in a low-channel network.

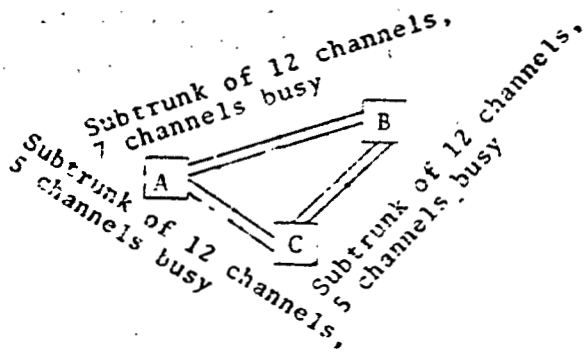


Figure 2.12. Multichannel Network with Undivided Subtrunks

The problem is that it is rare that precisely a standard group of 12 or 60 channels is required for communications. As a rule, the needs of the stations do not correspond to these numbers and will be either less than or greater than the standard subtrunk. For example (Figure 2.12), station A needs to organize communications with station B over 7 channels, with station C over 5 channels, while stations B and C need /49 5 channels between them. In order to satisfy the demands of station A, it must be allocated two subtrunks of 12 channels each, through one of which it transmits information over 7 channels to station B, while through the other it transmits information over 5 channels to station C. At the same time, one subtrunk each must be allocated to station B and C, through which they respond to station A. Furthermore, stations B and C each need another subtrunk for communications with each other over 5 channels.

Of the 72 simplex channels allocated for communications (6 subtrunks of 12 channels each), only 34 are actually in use, while the remaining 38 are free. This situation could not occur in a low channel communications network, but is typical for multichannel communications networks operating with unassigned and indivisible subtrunks.

However, due to its primary attribute -- the ability to serve many stations simultaneously -- a satellite communications system allows networks to be constructed significantly more economically. Suppose in this example station A radiates all information intended for stations B and C in a single 12 channel subtrunk (Figure 2.13),

received by both station B and C. Each of these stations separates its own channels from the subtrunk and filters out the others. In turn, both stations respond to station A and exchange information among themselves in their own 12 channel subtrunks. Then of the 36 simplex channels allocated in the network, 34 channels will be busy and only 2 will remain free. With other numbers of stations and other flows of information, the usage factor of the network channels may be either better or worse than in the example in question, but will always be significantly better than when indivisible subtrunks are used.

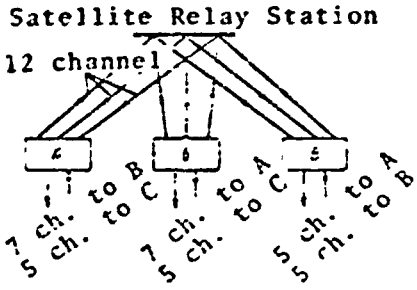


Figure 2.13. Multichannel Network with Divisible Subtrunks

Thus, the use of this method of network operation, based on the use of multiple-address subtrunks, produces a significant gain -- the necessary number of channels in the network and on the satellite is halved, while their load approaches the maximum possible.

However, operation in this manner requires that the ground stations have an excess of receiver equipment in comparison to their transmitting equipment, frequently quite a great excess. Furthermore, the channels addressed to each ground station must be separated at the ground station from each subtrunk, while the channels directed to other stations are filtered out.

Separation of channels from a large group with modern methods of frequency multiplexing in multichannel communications lines is possible only at the low frequency after complete breakdown of the group into individual channels. To do this, following the demultiplexing equipment, switches must be installed (Figure 2.14) to connect the channels addressed to the station in question to the channels on the lines approaching the station from the surface communications network. /50

When this principle of organization of communications is used, it is impossible to pair up simplex communications channels in advance. Each station can address channels in its subtrunk to any other station which receives the subtrunk. As the requirements change, the same channel may be addressed to different stations while the subtrunk itself is also assigned to the station only for a definite period of time, after which the station may operate in the same direction using another subtrunk.

The impossibility of pairing simplex channels makes the operation of a multichannel communications network more difficult. Each time a connection is established between two subscribers, it is necessary to organize not one duplex channel, as is done in

multichannel networks with centralized control (or in surface lines), but rather two independent simplex channels from subscriber A to subscriber B and back. This method of operation is possible only when there is a special control system in the network, establishing the connections required to organize a channel. Thus, the multichannel satellite communications network must be a single combined system with common network control of the operation of its channels.

Let us study the two main versions of construction of a multichannel communications network operating with unassigned, switchable channels.

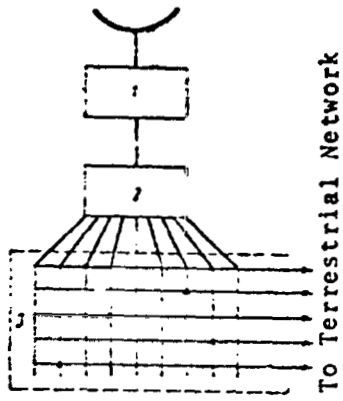


Figure 2.14. Switching System for Separation of Incoming Channels at Station of Multichannel Network: 1, Subtrunk Receiver; 2, (Receiving) Multiplex Equipment; 3, Coordinate Switch

The implest version of the operation of a multichannel network is the routing of communications through a center and centralized control. Such a network, as was shown above, can be expediently constructed in those cases when radial flows predominate and satellites are used which fly at moderate or low altitudes. The network has at least 2(S - 1) subtrunks. The central station radiates signals in at least (S - 1) subtrunks. If the streams of the individual peripheral stations are not added into a single standard subtrunk, the network can use more than (S - 1) subtrunks. Each peripheral station transmits information to all its correspondents (including the CS) on a common subtrunk and receives response information from the CS also on a common subtrunk. /51

Depending on the requirements of the subscribers, the streams of information exchanged by the stations change continually. The situation may arise that certain PS, for example A and B, require a large number of channels and occupy several subtrunks each at a given moment. At the same time, stations C and D may have very light traffic, and the CS radiates information to them in a single, common subtrunk. At the next moment, the situation changes, and the maximum flow shifts to C and D or some other stations. Thus, dynamic redistribution of information flow occurs within the network, and the channels are correspondingly redistributed between CS and PS.

The initiative in redistributing subtrunks in this network belongs to the central station, since it is this station which determines which PS need to have more subtrunks or, conversely, less subtrunks. The decision at the CS is made depending on the

requests of the stations. In some cases, a request for increased volume of communications may come from some individual PS, for example A, where the necessity of organizing a large number of channels with a number of other stations in the network -- CS and PS -- has arisen. However, most frequently the need to increase the number of lines between CS and station A will be detected at the CS, where the channels of many stations desiring to call station A are collected for relay.

Furthermore, since the CS switches the subtrunks, the addresses of all channels organized must be received and decoded at the CS.

It is also true that the main flow in the network, as was noted above, is concentrated in radial directions. Consequently, the CS, when a communications network is organized with relay through the center, must participate in the establishment of each new connection. Under these conditions, it seems most intelligent to use the principle of centralized control of the network.

The algorithm for establishment of connections in a multichannel satellite communications network with centralized control is as follows: if an order is received at peripheral station A from the surface communications network with which station A is connected for establishment of a connection with another PS, say B, the station seeks out a free channel in the subtrunk (radiated from the PS) for the call, then sends a call to the CS.

At the CS, the call is sent to the control apparatus -- the central control device (CCD). The CCD determines whether there are free channels in the subtrunk (or subtrunks) used for transmission to station B. If there is such a channel available, the CCS produces instructions for the central switch of the CS and for development of the required connections (a channel from subtrunk A is switched to a channel in the subtrunk radiated by the CS to A). At the same time, the CCS sends out the necessary instructions to the PS control devices and thereby makes the necessary connections in their switches. /52

If a channel must be organized between a PS and the CS, the sequence of operations is approximately the same, but in this case the CCS connects the PS channel to the matching apparatus and further to the surface communications lines and seeks out a free channel only in the subtrunks with which the CS is connected to the peripheral station in question.

The central station of the network (Figure 2.15) is a very complex station. First of all, it contains receiving and transmitting apparatus for the multichannel subtrunks, designed to receive all the information from the PS and transmit response information. In order to separate the multichannel subtrunks received into individual channels, the receiving end of a multiplexing apparatus is connected to each subtrunk received. To exchange service information between the

PS and CS, each subtrunk must also have a service channel, which may be organized, for example, in the subtrunks. The receiving apparatus for service communications also connects to each subtrunk.

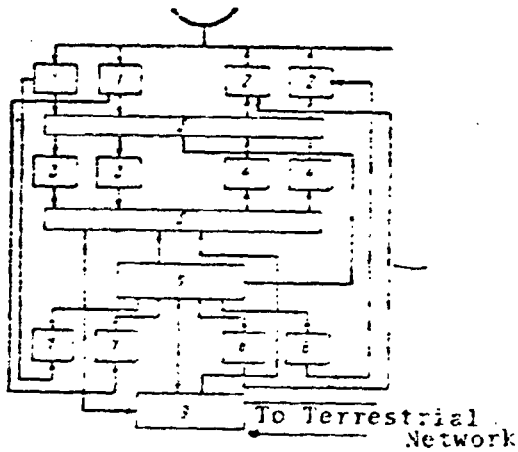


Figure 2.15. Central Station of a Multichannel Satellite Communications Network with Centralized Control: 1, Subtrunk Receiver; 2, Subtrunk Transmitter; 3, Receiving End of Multiplexing Apparatus; 4, Transmitting End of Multiplexing Apparatus; 5, Switching System; 6, Control Apparatus; 7, Service Communications Receiver; 8, Service Communications Transmitter; 9, Satellite and Surface Channel Matching Equipment

The individual channels produced at the output of the receiving ends of the multiplexing apparatus are transmitted to the switching system. There, according to the addresses, they are switched into new groups which must be transmitted to the peripheral stations. Some of the channels are sent by the switch to devices which connect the satellite channels to the surface lines connecting the subscribers to the satellite network. In turn, the information to be transmitted from CS to PS is fed to the switch and connected by it into groups formed for transmission to the radiating subtrunks. These groups are transmitted from the output of the switch system to the transmitting ends of the multiplexing apparatus and then to the subtrunk transmitters. The service channel signals are sent to the same transmitters. /53

The switching system makes the necessary connections on instructions received from the control apparatus. This apparatus receives from all PS and from the matching apparatus of the CS channels the information necessary for establishment of connection, which comes in over the service lines: calls, addresses of channels, etc. According to this information and the situation in the network, which is recorded in the memory of the control apparatus, instructions are developed for the CS switching system, as well as instructions for control of the PS switches; these instructions are sent over the service lines to the required PS.

The functional diagram of a peripheral station is similar to the diagram of the CS (Figure 2.16). The primary difference between PS and CS is the fact that the volume of receiving and transmitting apparatus at the PS is significantly less, as well as

the very important fact that the functions of the control and switching equipment at the PS are simpler than at the CS. The PS commutator connects the satellite channels over which information is received and transmitted with the CS to the channels of surface communications lines. The PS control apparatus develops the corresponding instructions for the switching system, based on instructions, orders and addresses received both from the CS and from the surface lines (from the matching apparatus).

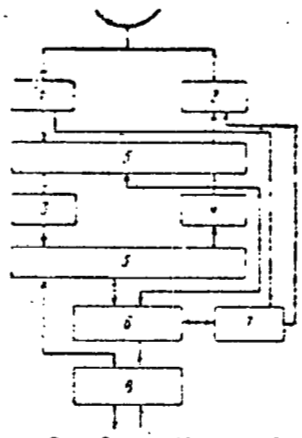


Figure 2.16. Peripheral Station of Multichannel Network with Centralized Control: 1, Receiver; 2, Transmitter; 3, Receiving End of Multiplexing Apparatus; 4, Transmitting End of Multiplexing Apparatus; 5, Switching System; 6, Control Apparatus; 7, Service Communications Transceiver Apparatus; 8, Apparatus for Matching to Surface Lines

In addition to control of channels and participation in the establishment of connections, the control center must perform the task of distributing subtrunks among stations. Quite obviously, since the information flow of PS changes with time, each PS requires various numbers of subtrunks at different times. At the HML, the number of subtrunks utilized by a PS reaches its maximum; during the periods of minimum load, a peripheral station needs only one subtrunk, and sometimes a PS may not need any satellite communications channels, since the entire load on the point where the PS is located is transmitted over surface communications lines. /54 During this time, the PS need only continue receiving the service communications channel from the central station.

Thus, we see that the control of subtrunks, their redistribution among stations of the network, is an important task of the CC. This task is performed by the central control device, and the subtrunk switches located both at the CS and at the PS (Figures 2.15 and 2.16). The CCS makes the decision on the basis of the analysis of the situation in the network, and also considering the statistical mean schedule of operation of the network. Subtrunks, with otherwise equivalent conditions, are allocated primarily to the PS which must transmit either priority information of great importance, or where past experience or other data indicates that the flow of

information in the immediate future will tend to increase.

The multichannel satellite communications network just studied is comparatively simple, but has a number of significant shortcomings. First of all, we must recall that a network with organization of communications through a center makes poor use of the capacity of the satellite relay system, which is the most complex and expensive device in the communications system. Furthermore, when high-altitude satellites are used, all communications channels between the PS have excessive delay and are not suitable for good quality telephone communications. Yet another defect in this network is the presence of the complex control center, which participates in the process of establishment of each connection. If the network is extensive, the volume of operations which must be performed by the CCD becomes astronomical. Such cumbersome equipment, of course, would have low reliability and would therefore have to be fully redundant, which would still further increase the complexity and cost of an already very expensive item.

All of these peculiarities of a network with communications through a center and centralized control lead to a single conclusion -- this type of network is good when radial information flow predominates strongly, but unsuitable if communications between PS make up a significant fraction of the total traffic. In this case, better results are achieved by a network with direct connection.

§ 2.7. Multichannel Satellite Communications Network Constructed According to Combined Principle

Let us recall the combined principle of organization of a network (Figure 2.17). In this type of network, communications between PS and CS are constructed as in the version described above with work through the center. Each PS transmits information designated for the center and remaining PS in one or more subtrunks. The CS extracts its own information and radiates response information to the PS in subtrunks assigned to each PS, so that any PS receives from the CS only the minimum number of subtrunks (no more than are radiated from the PS to the CS) necessary for two-way communications. /55

Communications between peripheral stations are developed on the direct communications principle. PS's radiate information designed for all other PS's in one beam (one or more subtrunks), and receive all information radiated by these other PS's, separating channels with the proper address from it.

The only shortcoming of the network based on the combined version is the necessity of receiving everything radiated by the PS at the CS and all other PS's and, consequently, the need for large

volumes of equipment for reception in comparison with the volume of transmitting equipment. This reduces the effectiveness of utilization of receiving equipment at the CS and PS. However, special measures can be taken to improve $B_{r\ ps}$ and $B_{r\ cs}$. For example, if the radial information flows of the PS are so great that they can occupy one or more subtrunks completely, these subtrunks can be assigned to the CS alone, and the CS will not need to receive information from PS not intended for it in the subtrunks.

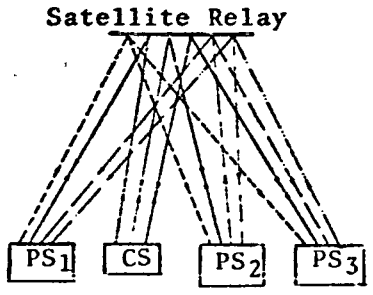


Figure 2.17. Organization of a Multichannel Satellite Network According to the Combined Principle

Thus, it is possible to reduce the necessary volume of receiving equipment at the CS and increase the factor $B_{r\ cs}$. An increase in $B_{r\ ps}$, i.e., a reduction in the volume of nonproductively used receiving equipment at the PS, can be achieved if we consider that the probability of simultaneous operation of the PS with all others is not great. As a rule, a PS at each moment in time operates with a limited number of correspondents. If we consider the demand statistics at each PS, we can significantly reduce the number of subtrunks received at the PS and, consequently, the volume of receiver apparatus. Furthermore, it may at times be suitable to use a limited number of subtrunks at a given PS, breaking off communications with some of the PS of the network (if this does not lead to significant increases in losses in the network as a whole), allowing the volume of receiving equipment at the PS to be significantly reduced.

We must note that the use of this type of measure to improve the economy of the network with communications through a center cannot have a significant effect, since in this case the losses in channels on the satellite, the double delay in PS-PS communications and the complexity of the CCD still remain. /56

This description of a network based on the combined principle and measures allowing the effectiveness of its utilization to be increased shows that the problem of redistribution of subtrunks between stations becomes quite important in this type of network. If PS-1 at a given moment is exchanging information with PS-2, PS-3 and PS-4, in the next moment it may need to organize communications with PS-5, while breaking off communications with PS-4. This means that PS-1 must stop receiving the subtrunk from PS-4 and begin receiving a subtrunk from PS-5. Additional subtrunks may also be

required at PS-1 for transmission, etc. The subtrunks in this type of network are continually dynamically redistributed. In solving the problem as to which PS should be given a subtrunk, it is necessary that:

1) the PS load the subtrunk to the maximum possible, i.e., that its load be greater than the load of competing PS;

2) the subtrunk, if it is given to a PS, can be received by all of the correspondents of the PS, i.e., that free equipment be available at the other PS for reception of the subtrunk;

3) it be considered whether the calling PS can receive the response information from all correspondents, i.e., whether it can receive the subtrunks of the PS with which communications are desired;

4) the importance of the information to be transmitted through the subtrunk by the PS be considered;

5) the tendency of flows of both competing PS to change be considered, and what the subtrunk be given to the PS, the flow of which, for one reason or another, will be greater and will continue for a longer time;

6) the influence of the assignment of the subtrunk to a given PS on the quality of servicing of the network be considered.

It can be stated in general that the problem of assigning or removing a subtrunk can be solved only by the network control center, which knows the situation in the network and can evaluate it properly, i.e., control of subtrunks in the network should obviously be centralized.

The network which we have described allows the stations to organize individual channels within the limits of the subtrunks assigned to them independently, bypassing the center, i.e., decentralized control can be used in organizing channels. In this case, each station of the network has switching and control equipment for making connections, this equipment being less complicated and smaller than the equipment for centralized control of channels would be.

Thus, in organizing a multichannel satellite communications network based on the combined principle, it is possible (and necessary) to construct a communications control system in the network which is also combined. The control of subtrunks, i.e., the control of flows of information and bundles of channels in the network, must be centralized, while the control of channels within subtrunks is decentralized. This organization of control allows construction of a flexible and reliable communications network. /57

Let us study the generalized algorithms for control of subtrunks and channels. Centralized control of subtrunks is performed approximately the same as centralized control of communications lines in low-channel communications networks. After

receiving a request for a subtrunk from a station, the control center (CC), evaluating the situation in the network as a whole and considering the conditions formulated above, either gives the instructions for switching of subtrunks or refuses the calling station.

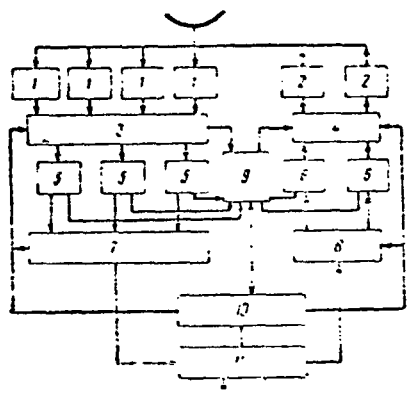
Subtrunks are switched by special switching systems located in the stations. Control of the operation of these switching systems is entrusted to the CC which, after making its decision, sends out instructions for the required connections directly to the subtrunk switching systems, where the necessary connections are made.

The algorithm for establishment of communications in telephone channels is more complex. After receiving a call signal from a surface communications line, the control device (CD) of the satellite station first determines whether there is a free channel in the subtrunk (or subtrunks) which the station is currently broadcasting. If there is such a channel, the CD determines whether the station can receive the other station toward which the call is directed, i.e., whether the calling station is receiving a subtrunk from the station being called. Let us assume that it is. It is then necessary to know whether the station called (in its subtrunk received by the calling station) has a free channel to answer the call, i.e., to organize duplex communications. In order to determine this, a request must be sent over the service line between the two stations to the station called. Suppose the station called responds that it has a free channel; then the calling station, which has received the request for communications from the subscriber, answers that a channel can be organized. If, however, it has found that there is no free channel either at the calling or at the called station, the subscriber is refused. The time necessary to determine whether it is possible to organize a channel may reach 600 msec, since the calling station must reach the station called through the satellite and receive an answer. To accelerate this procedure, for example, all stations may be required to mark their free channels. In this case, the calling station can rapidly see whether communications is possible with the station called, by checking to see whether there are channels in the subtrunk received by the calling station which are marked with the symbol of the station called.

After determining that it is possible to organize a channel and giving the corresponding response to the calling subscriber, the CD of the station begins organizing the channel. To do this, /58
the number of the free channels over which communications are to be conducted with the called station is noted, and the outgoing channel of the subscriber is connected to this channel, and at the same time the number of the channel is reported to the station called over the service line. At the called station, the number is determined of the channel over which it will respond to the calling station, and the two simplex channels are connected -- outgoing and incoming -- to the duplex channel of the surface line

to the called subscriber. The number of the channel over which information will be returned is transmitted to the calling station. The calling station, when it receives the number, makes the necessary connection of this channel to the incoming (simplex) channel of the calling subscriber. After this, the channel must be tested and then its readiness reported to the subscribers.

But as we see, the operations involved in establishing connections are rather complex, numerous and require some time. The time required to establish a connection, considering the switching time, performance of logic operations and signal propagation time over the service lines, reaches several seconds. However, this is not so bad, if we consider that the satellite channel replaces a long land line, which would have to pass through several switching centers, requiring at least the same amount of time.



To terrestrial network

Figure 2.18. Multichannel Satellite Communications Network Station Operating According to Combined Principle: 1, Subtrunk Receiver; 2, Subtrunk Transmitter; 3, Received Subtrunk Switching System; 4, Broadcast Subtrunk Switching System; 5, Receiving End of Multiplexing Apparatus; 6, Transmitting End of Multiplexing Apparatus; 7, Received Telephone Signal Switching System; 8, Transmitted Telephone Signal Switching System; 9, Service Communications Apparatus; 10, Control Device; 11, Matching Apparatus with Surface Channels

In the light of all the above, we can produce a block diagram of a satellite communications station for use in this network (Figure 2.18).

The station receives the signals of all subtrunks of the system; this is not too expensive, since the SHF receivers used are wideband units, and the cost of intermediate frequency sections and demodulators is low in comparison to the cost of the other equipment at the station. After receiving the group spectra, they are fed through the subtrunk switching system to the multiplexing units, the number of which determines the capabilities of the station for reception. This equipment is expensive, and its cost represents a significant fraction of the total cost of the station. The signals,

/59

broken down into individual channels, are fed to the telephone channel receiving switching system, where the channels addressed to the station are separated and switched to the proper channels of the land line matching apparatus, used to connect the subscribers to the satellite communications channels.

The transmitting portion of the station is similar to the receiving portion: signals from subscribers are fed through the telephone channel switching system to the multiplexing apparatus and then through the subtrunk switching system to the transmitter. Service communications in this network must be organized according to the "each with each" principle.

Figure 2.18 shows that the service communications transceiver separates its signals from the received subtrunks and includes them in the transmitted subtrunks through the multiplexing apparatus, although the service communications channels need not necessarily be combined with the main telephone channels: this is a problem in specific planning of equipment, and will be analyzed in other sections of the book.

The control apparatus of the station receives the information it requires from the surface line matching equipment and the service communications network from all other stations of the network. After performing the required logic operations, it outputs instructions to the switching system of the station and receives confirmation of performance of the instructions.

The central and peripheral stations in this network do not differ from each other in principle (with the exception of the volume of equipment and a few details); therefore, Figure 2.18 describes both CS and PS in general points.

REFERENCES

2.1. Karmazov, M. G., Organizatsiya i Planirovaniye Mezhdugorodnoy Telefonnoy Svyazi [Organization and Planning of Long-Distance Telephone Communications], Svyaz'izdat Press, 1960.

2.2. CCITT Documents:
Research Commission 11 -- Document No. 73, 1966.
Research Commission 13 -- Document No. 58, 1966.
Research Commission 11 -- Document No. 78, 1966.
Research Commission 16 -- Document No. 50, 1966.

CHAPTER 3. DESIGN OF SATELLITE COMMUNICATIONS NETWORKS

§ 3.1. Initial Data

The purpose of the design of a satellite communications network is determination of the necessary volumes of apparatus at ground stations and the capacity of the satellite relay station. Design is based on the required service quality for subscribers of the network, i.e., the permissible number of busy signals or the permissible waiting time. /60

Based on the calculated expediency of constructing satellite communications networks working with busy signals without waiting, we will study only the methods of designing these networks.

The initial data used for calculation consist of the traffic load on the network and the permissible losses in establishment of connections between subscribers of the network. The traffic can be fixed in various forms, but can be conveniently represented in the form illustrated by Table 3.1, showing the outgoing flow between stations of the network (in erlangs) in the hour of maximum load (HML).

TABLE 3.1

Stations of the Network	A	B	C	D	E
A		y_{AB}	y_{AC}	y_{AD}	y_{AE}
B	y_{BA}		y_{BC}	y_{BD}	y_{BE}
C	y_{CA}	y_{CB}		y_{CD}	y_{CE}
D	y_{DA}	y_{DB}	y_{DC}		y_{DE}
E	y_{EA}	y_{EB}	y_{EC}	y_{ED}	

Considering the capabilities of satellite communications to shift channels to the directions where they are most needed, Table 3.1 should be restructured in a single time scale, for example CS time, since these tables are usually drawn up regardless of time. The traffic tables should be constructed separately for each hour of the day. They will be similar to Table 3.1, but some /61

of the flows y_{ij} may equal zero, if no communications are required in a given hour in a given direction. Tables should also be made up for those hours where the flows are maximum, i.e., the hours (or hour) of maximum load of the network should be selected and all calculations based on it, considering that normal operation of the network will be assured during all other hours of the day.

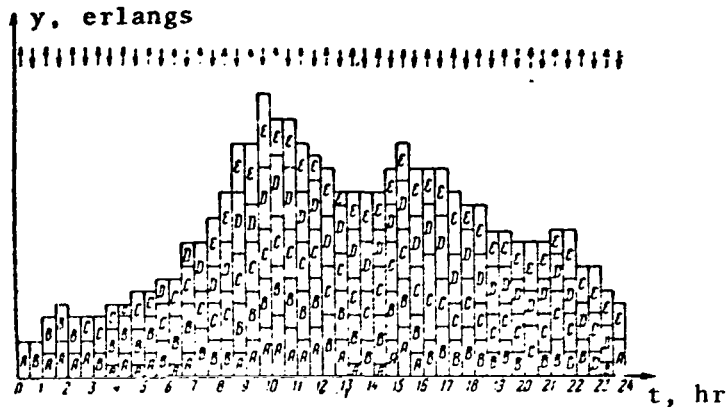


Figure 3.1. Traffic Load of Stations of the Network During the Day

If the stations in a planned network operate with significantly different local time (if the network is spread over a large east-west area), this restructuring of the initial table and selection of the hours of maximum load for the network may yield a significant savings in the necessary throughput capacity of the network and the satellite relay station.

After making up the traffic tables for each hour of the day (using the standard time for the network), the load of all stations in the network both as to incoming and as to outgoing communications must be determined for each hour of the day. This load is most conveniently represented for each station using graphs such as those illustrated on Figure 3.1. Here the abcissa shows the standard time of the network, the ordinate shows the load in erlangs. There are two columns for each hour: the first shows the outgoing load of the station for the hour, the second -- the incoming load. Each column consists of individual components making up the volume of the station in question. The load graphs for stations allow the necessary volume of equipment for the stations to be calculated. After constructing the load tables for the network and the load graphs for the stations, we can go on to design the network. /62

§ 3.2. Analysis of Versions of Utilization of Satellite Relay Station in the Organization of a Network

In order to determine the effectiveness of the use of the satellite relay, let us perform a comparative analysis of the versions described in § 2.2.

First of all, let us look at the version of direct communications of the stations with independent communications lines. We can write an expression for the capacity of the ith ground station of the network in channels Q_{gi}.

We will always write this capacity, unless otherwise stated, in simplex channels. In order to distinguish the channels of which we are speaking, channels over which information is transmitted or channels over which it is received, we will always include as a subscript the symbol "↑" if we are speaking of channels over which information is transmitted, and the symbol "↓" if we are speaking of channels over which information is received. Consequently, the inscription $Q_{\uparrow gi}$ represents the capacity in simplex channels of the group (bundle) of channels over which information is transmitted by ground station i, while the inscription $Q_{\downarrow gi}$ represents the same thing, but for the channels over which information is received.

The capacity $Q_{\uparrow gi}$ of the group of channels of station i, over which information is transmitted, will be equal to the sum of the capacities of the individual communications lines r_{ij} to all stations of the network in the version now in question, i.e.,

$$Q_{\uparrow gi} = r_{i1} + r_{i2} + r_{i3} + \ldots + r_{i(S-1)} = \sum_{j=1}^{S-1} r_{ij}, \tag{3.1}$$

$$Q_{\downarrow gi} = Q_{\uparrow gi}.$$

It is assumed in formula (3.1) that, studying the ith station, we perform addition for all remaining stations of the network, for which they are all assigned numbers from j = 1 to j = S - 1.

Let us agree that for each station studied, i.e., for each value of i, the stations in the network will be numbered so that the station in question is always assigned the last number, i.e., S, while the central station of the network (all communications networks have a central station) is always given the first number.

The total capacity of the communications network is determined by the expression

$$Q_{z0} = \sum_{i=1}^{S} Q_{zi} = \sum_{i=1}^{S} \sum_{j=1}^{S-1} r_{ij}.$$

the number of duplex communications lines in the network will be

$$l_d = \frac{S(S-1)}{2}.$$

Sometimes the network can be constructed so that all lines in the network have the identical number of channels $r = r_{ij}$. Its capacity is then defined by the expression /63

$$Q_{z0} = \sum_{i=1}^{S} l_i r.$$

here l_i is the number of lines going out from each station. One particular case of this type of system is a network which uses only single-channel communications lines. The capacity of the network is then simply equal to the number of lines, i.e., $Q_z = l_d$ duplex channels.

The throughput capacity of a satellite relay station is also most conveniently expressed in simplex lines or channels, since the power (or electronic) throughput capacity of the relay station is also generally defined in simplex channels.

In our case, the capacity of the relay station must be

$$Q_{r0} = \sum_{i=1}^{S} r_i.$$

and with lines with identical numbers of channels

$$Q_{r0} = 2l_d r,$$

The ground stations of the network must be designed for throughput capacity at HML, with identical capacity of receiving and transmitting equipment (receivers, transmitters and multiplexing apparatus):

$$Q_{g\ r0} = Q_{g\ tr0}.$$

Let us now study the extent to which the values of necessary equipment capacity at ground stations and on board the satellite change in the version with direct communication of stations using combined bundles of channels.

The capacity (total) of a subtrunk radiated by station i with this organization of the network is:

$$Q_{ti}=\sum_{j=1}^{S-1} r_{ij} \quad \text{channels},$$

where r_{ij} is the number of channels in the subtrunk of station i, addressed to station j.

The entire capacity S of the stations of the network, i.e., the necessary capacity of the relay station with direct communications of stations, is:

$$Q_{rd} = \sum_{i=1}^{S} Q_{ti}=\sum_{i=1}^{S}\sum_{j=1}^{S-1} r_{ij}. \qquad (3.2)$$

The capacity of a subtrunk determines the necessary volume of transmitting equipment at station i, i.e., /64

$$Q_{tr\ i} = Q_{ti} \text{ channels.}$$

The receiving equipment at station i should have

$$Q_{r\ i} = Q_r - Q_{ti}=\sum_{i=1}^{S}\sum_{j=1}^{S-1} r_{ij}-\sum_{j=1}^{S-1} r_{ij} \quad \text{channels},$$

although only Q_i channels are actually used in communications.

In order to allow the losses in utilization of equipment resulting from the method of organization of the network to be described, let us introduce a special utilization effectiveness factor B, the ratio of the volume of apparatus (in channels) actively participating in communications Q_a, to the volume which is potentially necessary (at a station, satellite or in the

network) to support normal operation Q_p, i.e.,

$$B = Q_a/Q_p.$$

In the network in question, the effectiveness factor of utilization of transmitter equipment will be equal to 1, while the effectiveness factor of utilization of receiver equipment

$$B_{r\,i} = \frac{Q_{+i}}{Q_{\cdot i}} = \frac{\sum_{j=1}^{S-1} r_{ij}}{\sum_{i=1}^{S}\sum_{j=1}^{S-1} r_{ij} - \sum_{j=1}^{S-1} r_{ij}},$$

i.e., is significantly less than 1.

The satellite relay station is used effectively with this method of operation, since all the channels which it relays are active. Consequently, $B_{rd} = 1$.

The situation is different in the version of communications with relay at the center. The use of the satellite in this version is worse. Actually, let us study the capacity of the relay station required in this network. Each of the (S - 1) PS radiates a signal to the satellite consisting of a subtrunk with a capacity of r_i channels. The CS, responding to the PS, also radiates (S - 1) signals in subtrunks with the capacity of r_i each. Consequently, the satellite must relay $n = 2(S - 1)$ signals, the total capacity of which is

$$Q_{rc} = 2\sum_{i=1}^{S-1} r_i = 2\sum_{i=1}^{S-1}\sum_{j=1}^{S-1} r_{ij} \quad \text{channels},$$

whereas direct communications in the network with the same flow would require that the satellite relay only

$$Q_{rd} = \sum_{i=1}^{S}\sum_{j=1}^{S-1} r_{ij} < Q_{rc} \text{ channels}.$$

The losses with communications through the center result from the fact that the communications signals between any two peripheral stations pass through the satellite twice -- once on the way from PS_i to the CS and once again on the way from the CS to PS_j. /65

We will characterize the use of the satellite by the relay station utilization effectiveness factor. For a network with communications through the center, we have

$$B_{rc} = \frac{Q_a}{Q_p} = \frac{Q_{rd}}{Q_{rc}} = \frac{\sum_{i=1}^{S}\sum_{j=1}^{S-1} r_{ij}}{2\sum_{i=1}^{S-1}\sum_{j=1}^{S-1} r_{ij}}.$$

In order to represent this factor more clearly, let us separate the volumes of communication with the central station from the expressions for Q_{rd} and Q_{rc}. For this we represent, as we noted above, the central station with the number 1. All peripheral stations can be numbered in any sequence with the exception of the particular station in question, the number of which will be S.

Each of the peripheral stations transmits to the other (S - 2) PS a volume of $r_{\downarrow ips} = \sum_{j=2}^{S-1} r_{ij}$ channels, $r_{\downarrow ics} = r_{il}$ channels to the central station.

With direct communications, as with communications through the center, one of the stations of the network may be central, in spite of the fact that it works in the same mode as the peripheral stations (generally, the flow of the CS is significantly greater than the flow of the PS). Thus, with communications through a center, the total number of channels on the satellite will be

$$Q_{rc} = 2\sum_{i=2}^{S} r_i\,pc + 2r_i\,tc = 2\sum_{i=2}^{S}\sum_{j=1}^{S-1} r_{ij} + 2\sum_{i=2}^{S} r_{in}.$$

In this expression, the flows are doubled, since everything radiated by the PS is relayed once more by the central station.

With direct communications, the total capacity of the satellite will be

$$Q_{rd} = \sum_{i=2}^{S} r_{i\,ps} + 2r_{i\,ts} = \sum_{i=2}^{S}\sum_{j=2}^{S-1} r_{ij} - 2\sum_{i=2}^{S} r_{is}. \qquad (3.3)$$

We can now write an expression for the satellite relay utilization effectiveness factor:

$$B_{rc} = \frac{\sum_{i=2}^{S}\sum_{j=2}^{S-1} r_{ij} - 2\sum_{i=2}^{S} r_{is}}{2\sum_{i=2}^{S}\sum_{j=2}^{S-1} r_{ij} - 2\sum_{i=2}^{S} r_{is}}. \qquad (3.4)$$

It is clear from expression (3.4) that the relay station utilization effectiveness factor in a network with communications through the center is less than 1, the difference between this factor and 1 being greater, the greater the share of traffic between peripheral stations in the total traffic of the network. The effectiveness of utilization of equipment at the central station is determined by the fact that, in addition to servicing its own communications with PS, it must receive and transmit, as well as switch, the communications channels connecting the peripheral stations with each other. This effectiveness can be represented by the factor $B_{cs\ c}$, the ratio of the number of channels used to receive (or transmit) information from PS to CS to the total number of channels required to service the central station, i.e., /66

$$B_{cs\ c} = \frac{Q_{\downarrow ps\ cs}}{Q_{\downarrow cs}} = \frac{\sum_{i=2}^{S} r_{is}}{\sum_{i=2}^{S} r_{is} + \sum_{i=2}^{S} r_{i\,ps}} = \frac{\sum_{i=2}^{S} r_{is}}{\sum_{i=2}^{S} r_{is} + \sum_{i=2}^{S}\sum_{j=2}^{S-1} r_{ij}}.$$

$B_{cs\ c}$ will be significantly less than 1 which, like the low satellite utilization effectiveness factor, is a shortcoming of this version.

Let us now analyze the combined version of organization of communications network.

Let us calculate the effectiveness of utilization of equipment in this type of network. First of all, we turn to the total capacity of the network, i.e., the necessary capacity of the relay station. Each ith peripheral station transmits a stream of information over $r_{ics} = r_{il}$ channels to the central station, and

transmits information to all peripheral stations over

$$r_{i\ ps} = \sum_{j=2}^{S-1} r_{ij} \text{ channels.}$$

All PS together radiate information to the CS over

$$r_{ps\ cs} = \sum_{i=2}^{S} r_{i1} \text{ channels}$$

and the CS requires the same number of channels to answer the PS, i.e.,

$$r_{cs\ ps} = r_{ps\ cs}.$$

The total flow of information radiated to the satellite in the combined version will be

$$Q_{rc} = 2r_{tps\ cs} + \sum_{i=2}^{S} r_{i1\ ps} = 2\sum_{i=1}^{S} r_{ti1} + \sum_{i=2}^{S}\sum_{j=2}^{S-1} r_{tij}.$$

Comparing the expression produced with (3.3), we see that $Q_{rc} = Q_{rd}$, i.e., the relay utilization effectiveness factor in the combined version /67

$$B_{rc} = Q_{rn}/Q_{rc} = 1.$$

The transmitting equipment at ground stations in this version has no ineffectively used channels, i.e., here $B_{g\ tr\ c} = 1$. The CS receives all subtrunks transmitted by the PS, since the PS radiate all information to the CS in a common subtrunk with the information to the PS. Consequently, the CS must have receiving equipment for

$$Q_{pr\ cs} = r_{ps\ cs} + \sum_{i=2}^{S} r_{i\ ps} = \sum_{i=2}^{S} r_{i1} \sum_{i=2}^{S}\sum_{j=2}^{S-1} r_{ij} \text{ channels,}$$

and each PS must have receiving equipment for

$$Q_{pr\ ps_i} = r_{cs\ i} + \sum_{i=2}^{S-1} r_{i\,ps} + r_{ps\ cs} =$$

$$= r_{11} + \sum_{i=2}^{S-1} r_{1i} + \sum_{i=2}^{S-1} \sum_{j=2}^{S-1} r_{ij} \quad \text{channels.}$$

We can now determine the effectiveness factors for utilization of receiving equipment at the CS and PS.

At the CS

$$B_{pr\ cs\ c} = \frac{r_{\downarrow ps\ cs}}{Q_{pr\ cs}} = \frac{r_{\downarrow ps1}}{r_{ps1} + \sum_{i=2}^{S} r_{i\,ps}}, \qquad (3.5)$$

at the ith PS

$$B_{pric} = \frac{r_{\downarrow csi} + r_{\downarrow psi}}{Q_{pri}} = \frac{r_{1i} + r_{psi}}{r_{1i} + \sum_{i=2}^{S-1} r_{i\,ps} + r_{ps1}}. \qquad (3.6)$$

Expressions (3.5) and (3.6) show that the ground station receiving equipment utilization effectiveness with the combined version is poorer than with communications through the center, since $B_{grcsc} < 1$ and $B_{grpsc} < 1$. However, it is interesting to compare the values of receiving equipment utilization effectiveness factors to the similar values for the version of direct communications.

Based on the expressions produced earlier for r_{ips} and r_{ics}, we can determine the capacity transmitted from each peripheral station for the version with direct communications:

$$r_{11} = \sum_{i=2}^{S-1} r_{1i} + r_{11} = r_{11} + r_{1\,ps},$$

divided into the flow for all PS and the individual flow to the CS, /68
and the volume of receiver apparatus at each peripheral station, also divided into channels from PS and channels from the CS:

$$Q_{pri} = \sum_{i=1}^{S}\sum_{j=1}^{S-1} r_{ij} - \sum_{j=1}^{S-1} r_{ij} = r_{ps1} + r_{1ps} + \sum_{i=2}^{S-1} r_{i\,ps} = \sum_{i=2}^{S-1}\sum_{j=2}^{S-1} r_{ij} + \sum_{i=2}^{S-1} r_{i1} + \sum_{j=2}^{S} r_{1j}.$$

The peripheral ground station receiving equipment utilization effectiveness factor will be

$$B_{rid} = \frac{r_i}{Q_{pri}} = \frac{r_{1i} + r_{psi}}{r_{1ps} + r_{ps1} + \sum_{i=2}^{S-1} r_{i\,ps}}.$$

Comparing the value of B_{rid} produced with expression (3.6), we see that increasing term r_{1ps} in the denominator over r_{1i}, i.e., the increase resulting from the fact that the ith PS must receive the entire flow of information transmitted from the CS, addressed to all PS, causes the effectiveness of utilization of receiver equipment of the PS in the direct communications mode to be lower than in the combined version.

In both versions, the central station must receive information transmitted by all peripheral stations, i.e., the receiving equipment utilization effectiveness for the CS is the same in both versions.

Let us study an example of creation of a satellite communications network including S stations. Suppose all peripheral stations exchange identical volumes of information with each other over $r_{ij} = r(i, j = 2, 3, 4, \ldots)$ channels, and identical volumes of information with the central station, but over $r_{i1} = r_0$ channels. The total capacity of the network and the satellite in this case will be $Q_\Sigma = Q_r$ channels.

With direct communications of the stations of the network using independent communications lines, communicatons of all PS with the CS require $l_{s,\,cs} = 2(S - 1)$ simplex communications lines

of r_0 channels each, while communications of the peripheral stations among themselves require $l_{s.ps} = (S - 1)(S - 2)$ simplex lines of r channels each. Therefore, the entire network requires

$$l_{s\Sigma} = 2(S-1) + (S-1)(S-2) = S(S-1)$$

simplex communications lines with a total capacity of

$$Q_{\Sigma n} = 2(S-1)r_0 + (S-1)(S-2)r = (S-1)[2r_0 + (S-2)r] \quad \text{channels.}$$

If the communications needs between stations increase, the total capacity of the network will increase linearly, but if the the number of stations in the network must be increased, the growth of required network capacity will be nonlinear and very rapid, as is shown on Figure 3.2 for $r_0 = 20$ and $r = 10$.

The central station must have (S - 1) communications lines of r_0 channels for transmission and (S - 1) lines for reception. Each PS must have one line of capacity r_0 channels and (S - 2) lines of capacity r channels each for transmission and reception. The total number of lines required in the network increases nonlinearly with increasing numbers of stations (Figure 3.3), but with any number of stations the effectiveness of utilization of the relay station and ground equipment in this network is equal to one. /69

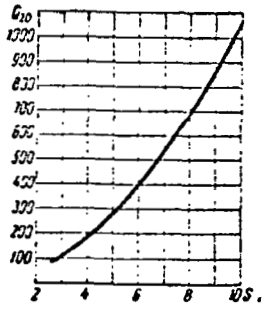


Figure 3.2. Capacity of Network as a Function of the Number of Stations for Communications Using Separate Lines

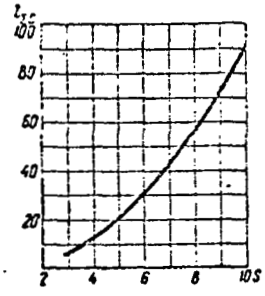


Figure 3.3. Required Number of Communications Lines as a Function of the Number of Stations in a Satellite Network with Individual Trunks

Let us now take the same network, but assume operation by direct communications using combined multichannel subtrunks. In this case, each PS must radiate a subtrunk with capacity

$$Q_{\uparrow i} = r_0 + (S - 2)r \text{ channels,}$$

and the central station must radiate a subtrunk with capacity

$$Q_{\uparrow 1} = (S - 1)r_0 \text{ channels.}$$

The total capacity of the network will be

$$Q_{\Sigma n} = Q_{\uparrow 1} + (S-1)Q_{\uparrow i} = (S-1)r_0 + (S-1)[r_0 + (S-2)r] = 2(S-1)r_0 + (S-1)(S-2)r \text{ channels, i.e.,}$$

equal to the capacity of the networks produced in the preceding version. As in the preceding version, the necessary number of channels in the network will increase (Figure 3.2) but with this growth, the number of subtrunks in the network remains as before, equal to the number of stations in the network.

It is easier to increase the capacity by increasing the number of stations in this network, since the necessary number of subtrunks increases linearly with an increasing number of stations in the network, whereas in the first version the number of communications lines increased significantly more rapidly than the number of stations.

A peripheral station must receive (S - 1) subtrunks with a total capacity of

$$Q_{\downarrow i} = (S-1)r_0 + (S-2)[r_0 + (S-2)r] \text{ channels,}$$

while the central station must receive (S - 1) subtrunks with a capacity of

$$Q_{\downarrow 1} = (S-1)[r_0 + (S-2)r] \text{ channels.}$$

The receiving equipment utilization factor at the ground stations in this case will be /70

$$B_{1rd} = \frac{(S-1)r_0}{(S-1)[r_0+(S-2)r]} = \frac{r_0}{r_0+(S-2)r} \quad \text{and}$$

$$B_{ird} = \frac{r_0+(S-2)r}{(S-1)r_0+(S-2)[r_0+(S-2)r]} = \frac{r_0+(S-2)r}{(S-1)r_0+(S-2)r_0+(S-2)^2 r}.$$

Figure 3.4 shows graphs of the change of B_{ir} as a function of the number of stations in the network where $r_0 = 20$ and $r = 10$ channels. We can see from the graphs that as the number of stations in the network increases, the receiving equipment utilization factor decreases, effectiveness B_{ir} being particularly low for the peripheral stations.

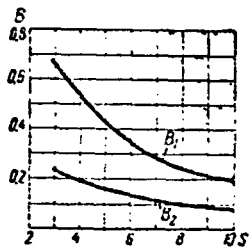


Figure 3.4. Receiving Equipment Utilization Effectiveness Factor at Central and Peripheral Stations in a Network with Direct Connection in Combined Subtrunks

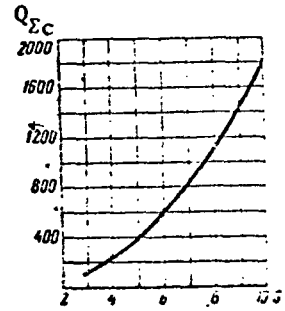


Figure 3.5. Necessary Network Capacity as a Function of Number of Stations in Network with Communications Organized through a Center

If communications is organized through a center in this same network, each PS then must transmit

$$Q_{1\,I} = r_0 + (S-2)r \qquad \text{channels,}$$

while the CS must transmit

$$Q_{I\,1} = (S-1)r_0 + (S-1)(S-2)r \qquad \text{channels.}$$

The total capacity of the network with this organization is

$$Q_{\Sigma c} = (S-1)r_0 + (S-1)(S-2)r + (S-1)[r_0 + (S-2)r] = 2(S-1)[r_0 + (S-2)r]$$

channels.

The function $Q_{\Sigma c} = f(S)$ is shown on Figure 3.5. As we can see, the necessary number of channels through the satellite for the same network studied above is significantly greater (compared with Figure 3.2), while the increase in the number of channels with increasing S is more rapid. All of this is explained by the properties of organization of communications through a center, which requires a double relay for all communications between PS. True, the effectiveness of the equipment at peripheral ground stations is equal to 1 in this case, which is an advantage in comparison with the version of direct communications with combined subtrunks studied above.

The use of the satellite will be characterized by the factor /71

$$B_{rc} = Q_{rn}/Q_{rc} = \frac{2(S-1)r_0 + (S-1)(S-2)r}{2(S-1)[r_0 + (S-2)r]} = \frac{2(S-1)r_0 + (S-1)(S-2)r}{2(S-1)r_0 + 2(S-1)(S-2)r} < 1$$

and will be less, the greater the share of communications between PS in comparison to communications of the PS with the center. If the PS are connected only to the center, i.e., if there are only radial connections in the network, $B_{rc} = 1$. Figure 3.6 shows a graph of $B_{rc} = f(S)$.

The effectiveness of utilization of central station equipment is determined by the formula

$$B_{1c} = \frac{(S-1)r_0}{(S-1)r_0 + (S-1)(S-2)} < 1$$

and will also decrease with increasing share of communications between PS.

Figure 3.7 shows the change in B_{rc} and B_{1c} as functions of growth of peripheral connections in the network, i.e., of the ratio r/r_0 with constant number of stations S = 10. As we can see, B_{rc} approaches 0.5 as r/r_0 increases, while B_{1c} approaches 0.

Figure 3.6. Relay Station Utilization Effectiveness Factor when Network is Organized with Communications through Center

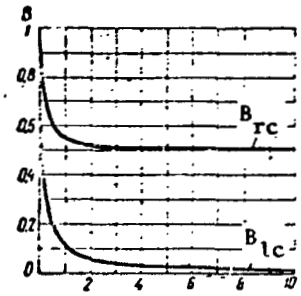


Figure 3.7. Equipment Utilization Effectiveness Factor for Relay Station and Central Station as Functions of Ratio r/r_1 with Organization of Communications through Center

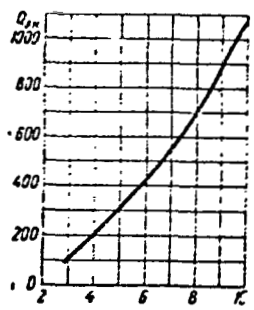


Figure 3.8. Necessary Capacity of Network and Relay Station as Functions of Number of Stations S with Combined Organization

Let us study, finally, the last version of organization of networks -- the combined version. In this case, the number of channels over which the CS transmits is

$$Q_{tcs} = (S - 1)r_0,$$

while the number of channels over which the PS transmit is

$$Q_{ti} = r_0 + (S - 2)r.$$

In all, consequently, the satellite receives

$$Q_{\Sigma c} = Q_{rc} = (S - 1)r_0 + (S - 1)[r_0 + (S - 2)r] \text{ channels.}$$

Figure 3.8 shows this dependence for $r_0 = 20$ and $r = 10$ channels.

The channel utilization effectiveness factor on the satellite is equal to 1, while the receiving equipment utilization effectiveness factor on the CS is /72

$$B_{r\ cs\ c} = \frac{(S-1)r_0}{(S-1)r_0+(S-1)(S-2)r}$$

and at the PS is

$$B_{r,\ i,\ c} = \frac{r_0+(S-2)r}{r_0+(S-2)r_0+(S-2)(S-2)r} = \frac{r_0+(S-2)r}{(S-1)r_0+(S-2)}.$$

Figure 3.9 shows these dependences as functions of the number of stations in the network.

We can see from Figures 3.8 and 3.9 that with the combined version of organization, the number of channels required in the network is not over the necessary number of channels in the version with direct connection (compare with Figure 3.2), while the utilization of receiver equipment at the PS is clearly better than in this version (compare with Figure 3.4).

Figure 3.10 shows these same factors $B_{r,\ 1,\ c}$ and $B_{r,\ i,\ c}$ as functions of the ratio between radial and peripheral connections r/r_0. For CS, the effectiveness of utilization of equipment is the same as in the version with communications through the center (see Figure 3.7), but at the PS the picture changes sharply. Here $B_{r,\ i,\ c}$ generally depends little on r/r_0, remaining equal to approximately 0.125.

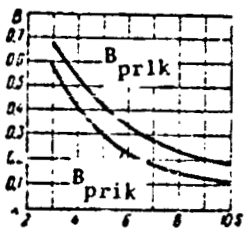


Figure 3.9. Ground Station Receiving Equipment Utilization Effectiveness Factor with Combined Organization as a Function of Number of Stations in the Network

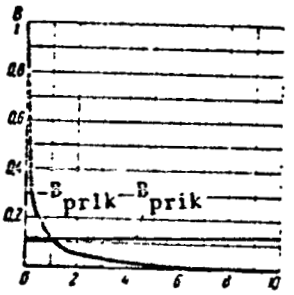


Figure 3.10. Ground Station Receiving Equipment Utilization Effectiveness Factor with Combined Version of Organization as a Function of Ratio Between Radial and Peripheral Stations r/r_0

Comparing the methods of utilization of satellites in organization of a network which we have studied, we can draw the following conclusions.

The first version -- using comparatively low-power communications lines, directly connecting the stations of the network -- is most economical as to effectiveness of utilization of equipment on the satellite and at the ground stations. However, it is complex and cannot provide a high throughput capacity of the satellite relay station, as will be shown in Chapter 8. This version

can be used in networks where there are comparatively large numbers of low-power stations, and the overall throughput capacity of the network need not be great. With this type of organization, the use of unassigned lines and channels, switched from one direction to the other as required, is quite significant. This operation can be particularly effective if the stations of the network are scattered over long east-west distances, i.e., if there are great differences in local time. /73

The second version -- direct connection of stations operating using powerful combined bundles of channels -- is suitable if there are comparatively few stations with more or less equal traffic, i.e., approximately equal flows of transmitted and received information. It is assumed in this case that high-altitude satellites are used.

If the network is planned for lower satellites and the delay in signal propagation with double relay is not too great, communications through a center should be considered, since it can be less expensive and more suitable than direct communications due to the sharp improvement of the effectiveness of ground station equipment utilization. A network with radial flow, of course, should be constructed according to the third version, with communications through a center, regardless of the altitude of the satellites.

If the network has predominantly radial flow toward the center (or toward several centers), but significant traffic between peripheral stations, depending on the altitude of the satellite, either the version of communications through a center (for low satellites) or the combined version (for high satellites) should be considered. Of course, in each specific case other, mixed versions of organization of the satellite communications network can be used.

The graphs on Figures 3.2-3.10 illustrate these conclusions rather well.

§ 3.3. Design of Centralized Low-Channel Network

A low-channel network with centralized control is characterized by the fact that any line to the satellite is a duplex line and is accessible to any subscriber, i.e., the network is serviced by a fully accessible bundle of duplex satellite channels. The subscriber in such a network, attempting to establish a connection through the satellite, may be refused due to lack of a free line on the satellite. The probability of a failure due to the satellite, i.e., of a loss, will be represented by p_s.

However, subscribers establish communications not directly, but rather through the ground stations of the network and reach the satellite communications network from outside, from the surface

network. For such a subscriber of the network, a ground station, for example GS, is a communications center, through which many subscribers in the zone serviced are connected. In attempting to establish a connection with their correspondents, subscribers may be refused either due to unavailability of free channels on the satellite, or due to unavailability of free connections at the transmitting GS (no free transmitters) or free inputs at the receiving GS (no free receivers). Thus, in addition to losses p_s, the line organized is subject to losses p_{gs_1} and p_{gs2}. /74

Let us determine the total losses in a satellite communications line, i.e., the probability of a busy signal either due to one of the stations or due to the satellite. When there are two or more independent sources of losses in a communications line (Figure 3.11), the total losses can be determined by the following simple arguments.

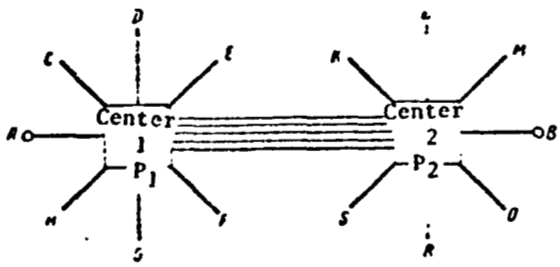


Figure 3.11. Network with Two Independent Switching Centers

Suppose there are two independent sources of losses in the line with failure probabilities p_1 and p_2. The total losses will be represented as p_Σ, the probability of achievement of a connection in the line as $1 - p_\Sigma = q$.

The probability of establishing a connection is equal to the probability of simultaneous occurrence of a number of events (probability that free channels will be found at both centers), i.e.,

$$q = (1-p_1)(1-p_2) = 1-p_1-p_2+p_1p_2.$$

Then the probability of failure to establish a connection will be

$$p_\Sigma = 1-q = p_1+p_2-p_1p_2.$$

As a rule, in communications lines the probability of failure is a small quantity, on the order of 1% or even a fraction of 1%. In these cases, the product p_1p_2 is a quantity two orders of magnitude less than p_1, p_2 or p_Σ, and this product can be ignored. Therefore, we can write for the case of addition of losses at two independent centers (sources of losses) with low values of losses that

$$p_\Sigma \approx p_1 + p_2.$$

Naturally, if there are more than two centers (sources of losses) in the line, but the losses at each one are independent of the others, we can write in general form /75

$$p_\Sigma \approx \sum_{i=1} p_i$$

under the condition that the absolute values of p_i and p_Σ are small.

Thus, if the losses arising at GS-1, GS-2 and the satellite are independent of each other, for the communications line in question we can write

$$p_1 = p_{gs1} + p_{gs2} + p_s. \qquad (3.7)$$

The independence of p_{gs} and p_s is quite obvious. Actually, in spite of the fact that, in determining losses, we study a certain communications line between two subscribers, the junctions included in the line, GS-1 and GS-2, as well as the satellite, operate in the network, i.e., service many communications lines and subscribers at once (Figure 3.11).

For example, a communications line can be organized through GS-1 to all possible stations of the network, through GS-2 as well. This means that if an output is found at GS-1 (a free transmitter), there may not be any free inputs, since all receivers of this group station may be occupied in communications with other stations of the network (not with GS-1). The call from GS-1 in this case is refused.

The correlation between cancellations at GS-1 and GS-2 will be great if all of the exchange of these stations, or at least most of it, is concentrated between them alone, i.e., when these stations are connected only with each other. In this case, if there is a free channel at GS-1 then, naturally, there will also be a free channel at GS-2.

If, however, GS-1 and GS-2 operate in a large network and each is connected to many stations, as shown on Figure 3.11, the share of communications between ther represents only a small fraction of the total exchange of traffic at each station and correlations between cancellations will approach zero.

This is even more true of a satellite servicing many GS. The probability p_s is quite independent of p_{gs1} or p_{gs2}.

Thus, we can state that if each station in a satellite communications network is connected to many others, expression (3.7) reflects rather precisely the losses in the communications line between the two subscribers of the network.

However, losses in establishment of connection between subscribers A and B are not limited to losses in the communications line p_1. During the process of establishment of a connection between A and B, group stations GS-1 and GS-2 contact the control center through the service communications lines, since only the CC /76 can find a line to the satellite and organize communications for them. However, in contacting the CC, a subscriber such as A may be refused because the service communications line is busy or the CC is busy processing the calls of other subscribers. Consequently, expression (3.7), defining losses in the network, should also include control losses p_c, consisting of the two components p_{cc} and p_{s1}, characterizing the losses due to cancellations by the CC and the service line. Naturally, these cancellations are independent of the other components of expression (3.7).

Thus, the overall losses in a low-channel satellite communications network with centralized control are defined by the expression:

$$p_\Sigma = p_1 + p_c = p_{gs1} + p_{gs2} + p_s + p_{cc} + p_{s1}. \qquad (3.8)$$

The value of p_Σ is usually fixed in planning the network at approximately 1-2%. The distribution of p_Σ into components according to expression (3.8) should be done individually in each case depending on the specific conditions. The way in which p_Σ is divided into components can be more precisely represented

after expressions are found for these components, i.e., the formulas for losses at the ground stations, satellite, control center and service lines.

Let us begin with determination of losses at the ground stations of the network. A group station servicing a zone, i.e., a large group of subscribers, has outgoing and incoming flows which can be represented by $Y_{i\ out}$ and $Y_{i\ in}$.

Let us emphasize once more that whether a flow is outgoing or incoming is determined by the call initiative, since in both cases the flows are served by duplex channels, i.e., for each channel regardless of whether it is an outgoing or incoming channel we must have both a transmitter and receiver at the station.

The outgoing flow is defined by calls originating in the zone of the surface communications network serviced by the given GS-1, and the satellite communications network directed toward any station. This flow is serviced, as at any communications center, by a bundle of equally accessible communications channels. The number of necessary channels in the bundle, if the number of subscribers served by the GS is large, is determined by the well-known Erlang formula on the basis of the permissible losses of calls [2.1]. The Erlang formula yields the probability of a loss p in a bundle of r channels if the incoming flow is equal to Y erlangs:

$$p = \frac{\frac{Y^r}{r!}}{\sum_{i=0}^{r} \frac{Y^i}{i!}}. \qquad (3.9)$$

This formula was used to compose tables and graphs which allow us to determine the necessary number of channels r from the fixed p and Y, losses p from the fixed Y and r, etc. Tables and graphs are presented in many books and handbooks on telephony and queueing theory [2.1], [3.1]. /77

Thus, for the outgoing flow $Y_{i\ out}$, using expression (3.9) or the tables, we can determine the necessary number of channels Q_{out} at a given GS. The number Q_{out} shows how many transmitters, receivers and terminal sets (in channels) must be available at a given station in order to transmit flow $Y_{i\ out}$ through the station with losses p_{gs}.

At the same time as the outgoing flow, the incoming flow $Y_{i\ in}$, determined by the calls arriving from other GS of the network to subscribers served by GS-1, also flows through station GS-1. On the average, as practical observations of flows in communications networks have shown, the incoming flow is approximately equal to the outgoing flow (or in any case has the same order of magnitude). Consequently, the number of incoming calls is approximately equal to the number of outgoing calls, i.e., for the incoming flow and for the outgoing flow, we can consider that the number of subscribers serviced by the station is rather high. In order to service flow $Y_{i\ in}$, the station requires a bundle of duplex equally accessible channels Q_{in}, which can also be defined using formula (3.9) or the tables for losses p_{gs2} (here we use p_{gs2}, since in the plan of a communications line described above, losses of incoming flow developed at GS-2).

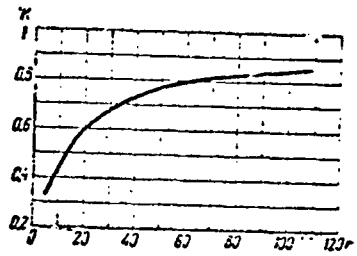


Figure 3.12. Channel Usage Factor as a Function of Trunk Capacity

In calculating the trunks Q_{out} and Q_{in}, we defined the necessary volume of transmitting-receiving and terminal apparatus (in duplex channels) for a given station. However, in a satellite communications network some additional savings in apparatus can be achieved by combining the incoming and outgoing flows of the station.

The reason is that according to the Erlang formula, the channel usage factor in a trunk increases as the number of channels in the trunk increases. This property is generally known and results from the statistical nature of flows of communications. Figure 3.12 shows the channel usage factor in a trunk k as a function of the number of channels in trunk r with constant losses p. As we can see from the figure, k increases significantly with increasing r.

For a satellite communications station, from the standpoint of calculation of the necessary number of channels and volume of equipment, the outgoing and incoming flows are in no way different. Actually, it is quite insignificant from where a call-request for a channel comes: from a subscriber of the station servicing the territory or from a subscriber associated with another station on the network. It is important only that with any call to the station it is necessary to find a free channel, i.e., a set of transceiving and terminal equipment capable of servicing the call. /78

Once this is found, the incoming and outgoing flows can be combined and the necessary number of channels at the station Q_i calculated from the summary flow $Y_i = Y_{i\ out} + Y_{i\ in}$. We will most likely achieve a gain in this case, since $Q_i < Q_{out} + Q_{in}$.

Flows $Y_{i\ out}$ and $Y_{i\ in}$ can be combined and a single bundle of servicing channels found only if the requests for losses in these bundles are identical. Therefore, in the following we will always consider

$$p_{rc_1} = p_{rc_2} = p_i.$$

Let us now turn to the relay station on the satellite. The total number of simplex channels to ground stations of the network, as follows from the above, is

$$Q_\Sigma = \sum_{i=1}^{s} Q_i.$$

If all channels worked continually, the satellite would also need Q_Σ channels. However, the operation of a network with unassigned channels is based on the use of the statistics of the flows; it is assumed that only a portion of the channels operate simultaneously in the network, i.e., that the stations are not simultaneously fully loaded. This means that the number of channels on the satellite can be less than the total number of channels of all the ground stations.

Furthermore, in a net with centralized control any channel on the satellite can be used by any station, using any set of apparatus at the station. This means that any channel of the satellite is accessible to any subscriber of the network, i.e., a satellite relay station with N_d channels is a fully accessible bundle of channels servicing all subscribers of the network. Under these conditions, we can calculate the bundle on the basis of the flow which will be transmitted through it. This flow is equal to the sum of the initial flows of all stations of the network, i.e.,

$$Y_s = \sum_{i=1}^{s} Y_{i\ out}. \qquad (3.10)$$

Since this flow is significantly greater than the flow of one station, this means that the use of channels on the satellite will be better than at the ground stations. For the satellite channels, the condition of application of the Erlang formula is also satisfied. In order to determine the necessary number of channels on the satellite, we must take the total flow Y_s from expression (3.10), fix the losses on the satellite p_s and, using the Erlang formula, find the number of channels on the satellite Q_r. /79

After determining the capacity of the relay station and the volume of equipment at the ground stations, let us study the losses resulting from operation of the control system and service communications lines. First of all, let us study the rate at which the control center should process calls arriving from subscribers.

It was demonstrated earlier that a satellite relay station services the total load of the system Y_s. From this it follows that the stream of calls arriving at the CC will be determined by this load Y_s, regardless of how it is distributed between the stations of the network. Thus, in order to determine the rate at which the CC should process incoming calls, we must find the rate of their arrival at the CC. The value of Y_s, as we know, is:

$$Y_s = \frac{\lambda}{\rho},$$

where λ is the intensity of the stream, i.e., the mean number of calls arriving at the system per unit time,

ρ is the servicing intensity, i.e., the mean number of calls serviced per unit time.

We usually know the value of $t_m = 1/\rho$, equal to the mean time of servicing of one call, i.e., the mean duration of one conversation. Consequently, it is easy to find the value of λ for a given flow Y_s. It is desirable that calls to the CC arrive in sequence and that it process the calls at a rate equal to the maximum rate of arrival of calls. This means that the servicing time of one call τ_0 must be selected so that not more than one call arrives at the CC during this time. Otherwise, the control device of the CC will not be able to process the second call which arrives and will generate a busy signal. At the same time, it should be noted that if a call is accepted for servicing by the CC, it will be processed according to the established conditions in a time of not greater than τ_0.

Since the flow of calls to the CC is a random flow, we can define τ_0 as the time during which not more than one call will arrive at the CC with fixed probability p_{call}. The probability of failure of the CC to service an arriving call will then be

$$p_{cc} = 1 - p_{call}.$$

In calculations related to telephone loads, it is usually assumed that the statistics of the flow of calls arriving for servicing follow the Poisson rule, according to which the probability of arrival of k calls in time t is determined by the formula /80

$$p_k(t) = \frac{(\lambda t)^k}{k} e^{-\lambda t} \qquad (3.11)$$

Using this formula, it is easy to calculate the required length of time τ_0 to assure that not more than one call will arrive in this time with probability p_{call}.

Actually, the probability that more than one call will arrive for servicing in time τ_0 is

$$p_{cc} = 1 - p_0(\tau_0) - p_1(\tau_0).$$

Here

$$p_0(\tau_0) = e^{-\lambda\tau_0}$$

is the probability of arrival of calls in time τ_0;

$$p_1(\tau_0) = \lambda\tau_0 e^{-\lambda\tau_0}$$

is the probability of arrival of precisely one call in time τ_0.

Thus,

$$p_{cc} = 1 - e^{-\lambda\tau_0} - \lambda t e^{-\lambda\tau_0} = 1 - (1 + \lambda\tau_0) e^{-\lambda\tau_0}.$$

With low $\lambda\tau_0$, we can expand expression $e^{-\lambda\tau 0}$ into a series and write

$$p_{cc} \approx 1-(1-\lambda\tau_0)(1-\lambda\tau_0)=(\lambda\tau_0)^2, \tag{3.12}$$

from which

$$\tau_0 = \frac{\sqrt{p_{cc}}}{\lambda}.$$

Thus, time interval τ_0 is, essentially, the minimum interval between two calls arriving at the CC, determined with probability $p_{call} = 1 - p_{cc}$. Since the time for servicing of one call at the CC is not over τ_0, with this probabiliy any call arriving will find the control device of the CC free and will be serviced, while the probability of a busy signal due to the CC is p_{cc}.

The operation of the service communications line is also determined by the interval of servicing of a call at the CC -- τ_0. In one second, not over $\theta_0 = 1/\tau_0$ calls can arrive at the CC. These calls are transmitted over the service communications lines between the CC and the stations of the network. These same lines also carry the response instructions of the CC to the stations following processing of calls. Thus, in one second, the service communications line must carry a maximum of $2\theta_0$ messages.

The number of calls (per second) from each station can be calculated from the load on this station, using expression (3.12). Using the flow of station i, let us determine λ_i, then the minimum interval between calls τ_i for fixed losses p_{sl} and, finally, the maximum number of calls per second $\theta_i = 1/\tau_i$. /81

Messages from the stations to the CC (calls) or answers of the CC to the stations (instructions) should not exceed τ_i in length. This is required in order to avoid a failure on the service line when a call arrives after τ_i sec due to its being busy transmitting the previous message. If we know the volume of one message -- call or instruction, which is determined in planning the control system -- we can determine the necessary throughput capacity of the service communications line from the ith station to the CC.

As an example, let us calculate the operation of the CC and service communications lines in the STAR system described above.

This system uses one common service communications line, occupied by all stations of the network and the control center in sequence. The operating speed of the line is 25,000 bps, the volume of one message is 80 bits. The time of transmission of one message is, consequently,

$$t_s = 80/25{,}000 = 3.2 \text{ msec.}$$

Considering that there must be time spaces between messages, resulting from the inaccuracy of the operation of the synchronization system, we will consider that 5 msec are set aside for each message[1]. Therefore, during one second the service communications line can carry 200 messages (100 from stations to the CC and 100 from the CC to the stations), i.e., the system can service up to 100 stations per second.

Occupying the service communications line in sequence, the stations automatically construct their signals for the CC into a sequence of calls. The CC can spend only 5 msec processing each call. Therefore, according to expression (3.12), if we assume $P_{CC} = 0.1\%$, the servicing rate at the CC for the load of the network is

$$\lambda = \frac{\sqrt{P_{CC}}}{\tau_0} \frac{\sqrt{0.001}}{5 \cdot 10^{-3}} = \frac{0.03}{5 \cdot 10^{-3}} = 6 \text{ calls/sec} = 21{,}600 \text{ calls/hr.}$$

Considering that the mean duration of a conversation is 5 minutes, the load serviced by the CC will be

$$Y_\Sigma = \frac{\lambda}{\rho} = \frac{21{,}600 \cdot 5}{60} = 1800 \text{ Erlangs}$$

If there are 100 identical stations in the system, each station can carry a load of $Y_i = 18$ erlangs and can transmit one call per second. The intensity of arrival of calls at the station will be

$$\lambda_i = \frac{18 \cdot 60}{5 \cdot 3600} = 0{,}06 \text{ calls/sec.}$$

[1] We could find no precise values in the literature.

Where τ_0 = 1 sec, the losses in service communications according to expression (3.12) will be

$$p_{s1} = (\lambda_i \tau_i)^2 = 0{,}06^2 = 0{,}0036 = 0{,}36\%.$$

The overall losses in the control system will be /82

$$p_c = p_{cc} + p_{s1} = 0.46\%.$$

It must be considered that there are no losses during the time the service line is occupied by the control center.

Let us attempt to estimate the permissible losses in the network, i.e., the values of components in expression (3.8).

The assigned value of losses in a planned device (satellite relay station, ground station, control center or service communications line) defines the volume, complexity and, consequently, cost of the device and of the entire communications system as a whole. This must be kept in mind in distributing losses p_Σ into components. Let us analyze communications systems from this standpoint.

The most complex and expensive devices in the communications systems in question are the satellite relay station and the transceiving and terminal equipment of the ground stations. In order to reduce the volume of this apparatus as far as possible, the primary share of losses must be accepted by this apparatus. The losses in the control system should be minimized, since the cost of this system represents a small portion of the total cost of the network, and some increase in cost of this equipment by the use of more rigid requirements for losses will in the final analysis be desirable, since it will allow the cost of the much more expensive and cumbersome ground transceiver and satellite relay equipment to be reduced. Therefore, the relationship $p_1 >> p_c$ should always be maintained.

Usually, losses in servicing a network with cancellations must be kept to 1-2%, i.e., p_Σ = 0.01-0.02. In this case we will arbitrarily consider that $p_1 = 10\ p_c$, i.e., p_1 = 0.009-0.018, while p_c = 0.001-0.002.

The value of $p_1 = p_s + 2p_i$ should probably be divided evenly between the satellite and two ground stations, i.e., it should be

considered that $p_s = p_i$ and $p_l = 3p_i$. With this solution, losses on the satellite will be less than the total losses of the ground stations, equal to $2p_i$, but it must be considered that the load on the satellite is significantly greater than the load on the ground stations and, consequently, even with identical losses, the use of channels on the satellite will be higher than at the ground stations. It is hardly necessary to shift a large portion of the losses p_l to the satellite, since this yields no significant gain in the load on the satellite relay, but significantly worsens the loading of the ground stations. If, on the other hand, the primary losses are shifted to the ground stations, this may lead to undesirable reduction in the load on the relay station.

The losses in the control system p_c can also be divided approximately evenly between the CC and the service communications line.

All of these figures are quite rough and are used only to make up the first version of loss distribution during planning. /83
This version should be calculated before the economic indicators are calculated, then refined during technical and economic optimization of the planned network considering the actual and prospective costs of equipment, satellite, etc.

§ 3.4. Network with Centralized Control and Direct Establishment of Connections Between Stations

From the point of view of design, this network is quite similar to that studied above.

Actually, with this organization of the network also, any channel on the satellite can be taken by any station of the network, i.e., the satellite relay station is in this case a fully accessible bundle of channels. The capacity of the relay station necessary for the operation of the network is determined from the total load which it must service. The losses p_s arising when the channels of the satellite are occupied by ground stations are independent of the load on these stations and are determined only by the relationship of the capacity of the satellite to the total flow passing through it. Calculation of that capacity and satellite losses can be performed using Erlang's formulas.

The ground stations should have a certain number of channels each in order to service the incoming load. The number of channels for any station, i.e., the number of transceiver and terminal equipment sets, can be determined from this load, by fixing losses p_i.

Just as in the network studied above with a control center organizing channels in the network in question, the losses at the calling station and called station will be independent of each other if the traffic between them is a small fraction of the total volume of each station.

Thus, the total losses in the communications line in a centralized network with direct establishment of connection will be determined by the expression already known

$$p_1 = p_s + p_{gs1} + p_{gs2} = p_s + 2p_i,$$

where the components of this expression are calculated by Erlang's formula individually for the ground stations and the satellite.

The difference between this network and that studied earlier is the operation of the control system and its service communications. There are no special service communications lines in this network, servicing the control system and, consequently, there is no loss component p_{sl}. However, at least two sources of losses are possible in the control system itself, which should be studied in detail.

With this version of organization and control, the ground station continuously observes the channels of the network, since calls may arrive at the station over any of these channels. The system of observation of the channels also searches for free channels in case of arrival of a call at the station from a subscriber of the surface network. The channel search process occupies a certain time interval τ_s and if another request for connection arrives during this interval, it (the second request) is refused, since the control system cannot service it. This means that the control system at the ground stations of the network creates losses, which will be represented by p_{ci}. The value of these losses can be determined by the ordinary methods used for solution of similar problems in telephony. /84

Since the flow of calls arriving at a given station is known and equal to Y_i, knowing the mean time that the channel is occupied by one call t_m, we can determine the intensity of the flow $\lambda_i = Y_i/t_m$, equal to the mean number of calls per unit time. Furthermore, knowing the mean time of servicing of one call in the station control system τ_s, we can find the load which must be serviced by the control system:

$$Y_{ic} = \lambda\tau_p.$$

C-2

The flow of calls arriving at the station, as we have already stated, will be considered simple. The servicing time of each call in the control system can be considered random with exponential distribution. The losses in the control system under these assumptions can be determined from Erlang's formula

$$p_{ci} = \frac{\frac{(Y_{ic})^k}{k!}}{\sum_{j=0}^{k} \frac{Y_{ic}}{j!}}$$

where k is the number of devices servicing incoming calls.

The second source of losses in the control system is characteristic only for satellite communications networks. The reason for these losses consists in the long propagation time of signals between two stations.

Let us imagine that station A has received a request from a subscriber of this station for establishment of a connection with some other station of the network. As soon as this request arrives, the control device of station A finds a free channel of the network, let us say the channel with number ξ, and sends a signal to occupy this channel to the control center. The busy signal reaches the CC after a time determined by the distance from station A to the satellite and from the satellite to the CC. In the worst case, for example when stationary satellites are used in the communications network, this time may reach 300 msec. After the busy signal from station A arrives at the CC, the control device of the CC removes the marker from the channel and /85
sends station A a signal permitting it to take the channel. However, the marker, after it is removed at the CC, will still be received by all other stations of the network for about 300 msec.

Even if we ignore the time necessary to process the call at station A and the CC, the stations of the network will receive the information that the channel is busy only 600 msec after it is actually taken by station A. Naturally, calls may also appear at the other stations of the network during this time and the free channel search system may stop at channel ξ at several of these stations. Taking channel ζ, another station of the network (for example, B) will also send the busy signal to the CC, but will be refused.

Let us study how the losses of this type P_{cb} can be calculated.

It is simplest to find the control losses arising in the process of taking a free channel in a network, where the planning calls for operation with one free marked channel. In this case, all stations at which requests for a channel appear will attempt to occupy the same free channel, the one marked at the given moment. If two or more calls appear in the network during the course of 600 msec, this will cause a failure. The probability of a failure in this case is equal to the probability of appearance of more than one call in the network during the delay time τ_d = 600 msec.

Considering that the flow of requests for channels arriving at the CC from all stations of the network is simple, it is easy to determine the probability of appearance of more than one call in time τ_d using formula (3.12):

$$p_{cb} = (\lambda_\Sigma \tau_d)^2 = 0.36\lambda_\Sigma^2, \qquad (3.13)$$

where λ_Σ is the intensity of appearance of calls with total load on the entire network Y_Σ.

For example, if the total load in the network is 124 erlangs, as in the example studied in § 2.1, then

$$\lambda_\Sigma = \frac{124 \cdot 60}{5 \cdot 3600} = 0.4 \text{ calls/sec}$$

and the losses due to simultaneous occupation of the one free channel by different stations will be

$$p_{cb} = 0.36 \cdot 0.16 = 5.75\%.$$

As we can see, the losses are too great and, consequently, this network should not use the method of marking one free channel.

If the control system of the network marks several (or all) free channels, losses p_{cb} will be significantly less, since the probability of a failure in this case will correspond to the probability that at least two stations of the network will attempt to occupy the same channel during time τ_d, even though each station could take any one of the free channels. /86

If station A takes free marked channel ξ, the probability that it will be refused p_{cb} will be equal to the probability that

at least one other station has received a call and taken channel ξ during the preceding time interval $\tau_d \leq 600$ msec, i.e.,

$$p_{cb} = \sum_{i=1}^{s-1} p_{call\ i} p_{\xi i},$$

where $p_{call\ i}$ is the probability of appearance of at least one call at the ith station in time interval $\tau_d \leq 600$ msec;
$p_{\xi i}$ is the probability that the ith station has taken channel ξ to service this call; $p_{call\ i}$ and $p_{\xi i}$ are mutually independent.

Probability $p_{call\ i}$ is calculated for a simple flow using Poisson's formula and is equal to

$$p_{call\ i} = 1 - p_0(\tau_g) = 1 - e^{-\lambda_i \tau_g} \approx 1 - 1 + \lambda_i \tau_g = \lambda_i \tau_g,$$

where λ_i is the intensity of the flow of calls at station i.

After receiving a call, station i takes any free marked channel. The probability that this will be channel ξ is equal to $p_\xi = 1/Q_{rf}$, where Q_{rf} is the number of free channels on the satellite relay station. The number of free channels can be found by using the expression, known from queueing theory [3.2], for the mean number of busy channels

$$Q_{rb} = \frac{\sum_{k=1}^{Q_p} \frac{1}{(k-1)!} c^k}{\sum_{k=0}^{Q_p} \frac{1}{k!} c^k}. \qquad (3.14)$$

Thus,

$$p_\xi = 1/(Q_p - Q_{cb}) \text{ and } p_{cb} = \sum_{i=1}^{s-1} Q_p - Q_{rb}$$

where Q_{rb} is determined from formula (3.14)[1].

We can assume that the probability that any station on the network will occupy any given free channel is identical for all stations.

Furthermore, it is expedient to extend our addition in the expression for p_{cb} to S. /87

After this we can write the inequality

$$p_{cb} < p_{\xi} \sum_{i=1}^{s} p_{\text{call } i},$$

but $\sum_{i=1}^{s} p_{\text{call } i}$ is the probability of appearance of a call in the network in time τ_d and is equal, according to expression (3.13), to

$$\sum_{i=1}^{s} p_{\text{call } i} = (\lambda_{\Sigma}\tau_3)^2.$$

It follows from this that when all free channels in the network are marked, the probability of losses due to simultaneous taking of the same marked channel by two stations will be

$$p_{cb} < (\lambda_{\Sigma}\tau_3)^2 p_{\xi},$$

i.e., $1/p_{\xi}$ times less than in a network where only one free channel is marked.

§ 3.5. Design of a Low-Channel Communications Network with Decentralized Control

In this network, losses arise only at the ground stations, and independently for each one, as in a network with centralized control; consequently,

$$p_{gs} = p_{gs1} + p_{gs2}.$$

[1]Calculation by formula (3.14) should be done with the use of tables, which may be found for instance in [2.1].

In the satellite in this case there can be no losses, since any station contacts the satellite at a frequency assigned to it, regardless of how many signals are already being relayed by the satellite.

In a network with decentralized control, the satellite relay station can be looked upon as a system with an unlimited number of servicing devices, in contrast to the relay station in a network with centralized control, which has a limited number of channels and therefore introduces losses to the communications line between stations of the network.

The losses in the line for a network with decentralized control

$$p_l = 2p_{gs} = 2p_i,$$

or, as before, we can assume that

$$p_{gs1} = p_{gs2} = p_{gs} = p_i.$$

However, as was stated above in the description of a network with decentralized control, the satellite relay station in these networks can reradiate only a limited number of signals N_s, less than the number of addresses N_a, with the required quality. /88

The total number of channels on the ground stations in the network is N_a, but in order to assure normal quality of operation of the channels with a fixed signal/noise ratio, the maximum load on the network Y_Σ should be such that the number of channels operating simultaneously in the network will not be over N_s with a fixed probability p_{ch}. Probability p_{ch} is determined by the quality norms and is generally 0.998-0.999.

Thus, an important stage in design of a network with decentralized control is selection of the maximum permissible load on the network Y_Σ, which can be serviced by the relay station on the satellite.

The relay station can be looked upon as a system with an unlimited number of servicing devices (channels), Y_Σ can be determined from the expression for the probability of occupation of a given number of servicing devices k, known in queueing theory [3.2]:

$$p_k = (1/k!)y^k e^{-y}.$$

Any number of channels from 0 to N_a may be busy in the network simultaneously. The probability that the number of channels simultaneously busy will not be greater than χ is:

$$p_\chi = \sum_{k=0}^{\chi} \frac{1}{k!} y^k e^{-y}. \qquad (3.15)$$

In order to assure normal operating conditions of the network, we must find $y = Y_\Sigma$ from expression (3.15) under the condition that $p_\chi = p_{ch}$ and $\chi = N_s$, i.e., from the expression

$$p_{ch} = \sum_{k=0}^{N_s} \frac{1}{k!} Y_\Sigma^k e^{-Y_\Sigma}.$$

In our calculation, we should use the special tables, for example those printed in [3.3].

After defining Y_Σ, we can calculate the necessary capacities of the individual stations, always fulfilling the condition that the number of channels of all ground stations together is not greater than N_a, i.e.,

$$\sum_{i=1}^{S} Q_i \leqslant N_a. \qquad (3.16)$$

If we find that this condition is impossible, the network should not be constructed according to the decentralized principle. In this case, addresses would have to be assigned simultaneously to several channels, i.e., a network with incompletely accessible channels would have to be used and, consequently, additional losses would result. Calculation of the necessary number of channels at the ground stations should be performed using the Erlang formula with losses p_i /89

The losses in the control system in a network with decentralized control arise mainly due to the service communications lines. If two stations attempt to call a third station simultaneously, only one can take the service and call channel, the second will be refused.

The time the CSC is busy is determined by the time required to service one call τ_{call}. The probability of arrival of more than one call during time τ_{call} can be determined by formula (2.12) and is the probability of a failure, i.e.,

$$p_{ci} = (\lambda_i \tau_{call})^2,$$

where λ_i is the intensity of arrival of calls at the third station.

Thus, in order to estimate the losses due to the control system in the version of network organization which we are studying, we must determine p_{ci} for the station with maximum traffic and consider this the probability of equal probability of losses in the network, i.e., $p_{ci} = p_c$. In this case, the losses will actually be lower than p_c at all other stations with less traffic.

Comparing the three versions of organization of low-channel satellite communications networks which we have studied, we must note that the least losses are found in the network with decentralized control, since losses in this network arise only at the ground stations and in the CSC of each station. However, this network can be considered expedient only if condition (3.16) is fulfilled; we must consider that if in the future as the network develops the number of channels at the ground stations increases over N_a, incompletely accessible channels will have to be introduced in the network, which will increase the losses above the fixed limit.

A network with centralized control and direct establishment of a connection is convenient in those cases when the stations of the network do not require many channels in their operation, i.e., when the system of search for a free channel and reception of a call at the station is not too cumbersome and slow. These networks are expedient in those cases when the subscribers of the network operate in groups, using a limited number of satellite channels. For example, we can imagine a network servicing several departments, the stations of each department using their own specialized group of channels. In such a network, the total number of signals relayed by the satellite may be great, the stations of the network may be numerous and condition (3.16) of applicability of the version with decentralized control may not be fulfilled. At the same time, construction of a network with centralized control under these conditions is also not expedient, since each group must have its own center, the centers of different groups frequently being geographically separated. /90

A network with centralized control is obviously most expedient in those cases when the operation of a large number of stations must be supported, covering a large area, a country or even several countries, as in the STAR network.

§ 3.6. Design of a Multichannel Network with Switched Subtrunks (Groups of Channels)

In describing multichannel communications networks, we have shown that these networks can be constructed in three main versions -- with switched subtrunks, with communications through the center and the combined version. Let us study the methods of design of each of these.

The operation of a network with switched subtrunks is similar to the operation of a low-channel communications network with centralized control. The only difference is that the CC in the multichannel communications network allocates not single-channel, but rather multichannel communications lines to the stations -- subtrunks with capacities of 12, 60 or more channels.

Each pair of stations in the network is connected by one or more multichannel lines. As long as the capacity of these lines is satisfactory to service the load between the stations, the losses are determined only by the magnitude of this load and the number of channels in the line. Calculation of losses is similar to calculation of losses in a surface communications line between two stations or centers. Therefore, in this case the losses in the communications line are equal to the losses at one station, i.e., $p_l = p_i$.

If the stations of the network are not satisfied by the subtrunks allocated, they send a request to the CC to allocate additional lines. If there are no free subtrunks in the network, the CC refuses. This cancellation immediately sharply increases the losses in the network. Obviously, the communications network should be designed so that these failures will not arise. To do this, one should calculate the necessary number of channels and subtrunks in the network at the HML and select the capacity of the satellite relay system accordingly. Failures in the network due to unavailability of free subtrunks will not arise and the operation of the CC can be reduced to simple distribution of the subtrunks according to need, to prevent the stations from operating with underloaded subtrunks.

The losses in the control system of the network can be reduced to a negligible quantity, since the stations of the network request new subtrunks from the CC comparatively rarely and the probability that two calls will arrive at the same time is

negligible. Consequently, we can consider that $p_c = 0$, so that $p_\Sigma = p_1 = p_i$. /91

Thus, in designing a network with switched subtrunks, the necessary number of channels r between each pair of stations should be calculated for the HML, using the Erlang formula with losses $p_\Sigma = p_i$. After determining these capacities, the necessary number of subtrunks of the network n should be found, rounding the capacity of lines between stations to the nearest standard number of channels in the subtrunk -- 12, 60, etc.

The capacity of the satellite relay station will be equal to the number of subtrunks in the network n.

§ 3.7. Design of Multichannel Network with Communications Through a Center

Let us study the points on the line, connecting two peripheral stations of the network, where losses will arise. When a call appears at PS-1, the control device of this station seeks a free simplex channel in its subtrunk (or subtrunks), connecting PS-1 to the CS and a simplex response channel in the subtrunk from the CS to PS-1, since both PS-1 and CS know the subtrunks where operations are occurring. If the call is directed to another peripheral station -- PS-2 -- the CS performs the required switching and seeks a free simplex channel from the CS to PS-2 and a simplex channel from PS-2 tc the CS. PS-2 is connected with all stations of the network and, naturally, sometimes all channels in the CS-PS-2 line may be busy with communications of PS-2 with the CS, PS-3, PS-4, ..., PS-(S - 1). That is, regardless of the availability of a free channel in the PS-1-CS sector, there may be no free channel in the CS-PS-2 sector, and the call from PS-1 will be refused.

Thus, the losses p_1 in the communications line arise at three points in the network -- at the calling and called stations and at the CS, i.e., $p_1 = p_{cs} + p_i + p_j$, if communications are to be established between two PS. If communications are required between the PS and the CS, the losses may appear only at the CS and the PS and $p_1 = p_{cs} + p_i$.

Calculation of losses at each station should be performed using the Erlang formula, since in this case there is a simple flow of calls, arriving over a fully accessible bundle of channels. It should be emphasized that as a rule, the CS, in order to save channels, transmits information to several PS simultaneously in a common subtrunk, so that the number of simplex channels over which the CS operates toward the PS will be less than the total number of channels at the PS.

If the load between the CS and the PS in question is great and a bundle which is near the capacity of the bundle from the PS to the CS is actually in use for CS-PS communications, the losses at the PS and CS become dependent, since occupation of several channels in the direction from the PS to the CS causes simultaneous occupation of the same number of channels in the direction from the CS to the PS. When another new channel is organized between the PS and the CS, at the last moment there are no free channels, since all free channels in the CS bundle are already occupied in transmission of information to the PS. /92

Calculation of the losses in this type of PS-CS communications line represent significant difficulty, due to the fact of this correlation. However, the worst case, i.e., the case of maximum losses in the line, will occur when this correlation is absent and losses at the PS and CS are simply added.

There is no significant dependence between failures at the PS and CS if a single bundle of channels from the CS is used for simultaneous work with many PS, i.e., when the share of channels of the PS in question in the total number of channels in the bundle is slight.

In order to simplify calculations, we can limit the estimation of losses at the worst point, i.e., with total independence between losses at the PS and the CS. Then the losses in the line between the two PS will be

$$p_i = p_i + p_j + p_{cs1} + p_{cs2},$$

where p_{cs1} and p_{cs2} are the losses at the CS in the direction of PS-1 and PS-2 respectively.

Assuming $p_i = p_j$ and $p_{cs1} = p_{cs2}$, we produce:

$$p_{l\ ps\text{-}ps} = 2p_i + 2p_{cs}.$$

In the line between the PS and the CS

$$p_{l\ ps\text{-}cs} = p_i + p_{cs} = (1/2)p_{l\ ps\text{-}ps}.$$

The losses at the central station should be determined from the summary load for a bundle (subtrunk) of channels, servicing several PS simultaneously; losses at the PS are calculated only

for the load of this PS. Calculations utilize the Erlang formula.

In addition to the communications lines, losses, as with other methods of organization of the network, may also arise in the control system. The sources of losses here may be the control devices of the CS and the PS.

Actually, the number of calls arriving at the PS from the surface network at times may be so great that the CD at the PS cannot process them all. The same situation may arise at the CS. In order to avoid these losses, the CS at the PS and particularly at the CS must work fast. Furthermore, in some cases a waiting system may be introduced at the input of the CC, reducing the number of failures.

Since the control systems are complex, a slight additional complication allowing p_c to be reduced to a negligible quantity is always expedient. Based on this, devices of this type in a planned network should be designed to operate with losses on the order of p_c = 0.05-0.1%, then simply ignored as the network is designed. /93

Calculation of the necessary CD speed is performed by calculation of the speed of the CD of a control center in a low-channel centralized network, described above.

§ 3.8. Design of a Multichannel Network Organized According to the Combined Version and Using Combined Control

In designing this network, one must distinguish the information flows between the CS and peripheral stations from the flows between PS. Communications of the CS with the PS are conducted over special subtrunks, i.e., a bundle of channels consisting of one or more subtrunks services definite, assigned PS. Consequently, in calculating the necessary capacity of this bundle, one must take the total load on those PS which are to use the bundle of channels in question.

The larger the total bundle of channels with the CS servicing a group of PS, the greater the savings in the number of channels in the bundle which can be achieved by increasing its power. However, increasing the capacity of a bundle requires that a larger number of channels be received at the PS which are designed for other PS of the group, i.e., reduces the utilization factor of PS receiving apparatus. In planning a network, one should seek the optimal distribution of PS among the groups, for which the total cost of the network is minimal.

The bundle of channels from the PS may be common, i.e., directed both toward the CS and toward other PS, although in

certain cases (with heavy loads), a special bundle of channels may be allocated to the PS for communications with the CS. In the general case, the number of channels necessary at a given PS for communications with the CS and the remaining PS of the network is calculated from the load with these stations. Here, as before, we have in mind the total, i.e., incoming and outgoing, load.

Figure 3.13 shows a network of four points, connected to each other according to the principle of "each with each" by multichannel communications lines. Each station transmits its information to the other stations in the network through a common bundle of channels. It is convenient to study the operation of one station of the network, for example A, relative to all the other stations simultaneously, i.e., to divide the entire network into two zones: the first zone includes station A, the second zone -- all other stations. Station A transmits information into zone 2 through the common bundle of channels and extracts response information from the bundles of all PS included in this zone, i.e., station A is in communication with all of zone 2. When a call appears at station A, it (the call) may be refused if all channels of station A are busy, i.e., if a free channel in zone 2 cannot be found at station A. /94 However, even if such a channel is found, the connection will not necessarily be established, since station B of zone 2, toward which the call is directed, may have no free channels to answer the call, i.e., to organize a duplex channel. All the channels of station B may be busy in communications within the zone. Thus, in establishing communications from station A to any other station of the network, losses p_1 arise at two points on the line -- at the calling station and the called stations.

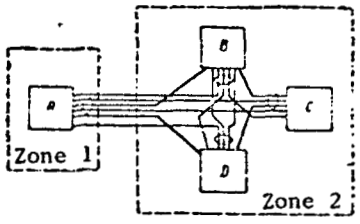


Figure 3.13. Division of Satellite Network into Two Zones

The total load on station A consists of a number of components, defining the communications of A with all the stations of zone 2, i.e.

$$Y_{A2} = Y_{AB} + y_{AC} + y_{AD}. \quad (3.17)$$

If each of the components in expression (3.17) is significantly less than the total load, i.e.

$$Y_{A2} >> y_{AB} \text{ or } y_{AC} \text{ or } y_{AD},$$

and if the same condition is maintained at all other stations of the network, we can consider that the losses at each station are

independent of the status of the other stations, i.e., we can consider the losses at individual stations of the network p_i independent. In this case, as before, the total losses in establishment of a connection in a line between two stations will be equal to the sum of the losses at each of them (with low absolute value of losses), i.e., we can write that $p_l = p_i + p_j = 2p_i$, since in this case it is expedient to assume the losses at all stations identical.

The losses in the control system of a network should be made, as in the preceding version of organization studied, negligible and, consequently, in our design of the network we need not consider them. Calculation of the CD of the control system should be performed as was done in studying low channel networks, i.e., according to the summary flow of calls which must be serviced by the control device. The capacity of the satellite relay system is designed for the maximum number of subtrunks used in the network during the HML, as is done in designing a network with communications through a center or a network with switched subtrunks.

REFERENCES

3.1. Markhay, Ye. V., V. N. Roginskiy, A. D. Kharkevich, Avtomaticheskaya telefoniya [Automatic Telephony], Svyaz'izdat Press, 1960.

3.2. Gnedenko, B. V., I. N. Kovalenko, Vvedeniye v teoriyu massovogo obsluzhivaniya [Introduction to Queueing Theory], Nauka Press, 1966.

3.3. Zakharov, G. L., N. P. Varakosin, Raschet kolichestva kanalov svyazi pri obsluzhivanii s ozhidaniyem [Calculation of the Number of Communications Channels Required for Queuing with Delay], Svyaz' Press, 1967.'.

CHAPTER 4. PECULIARITIES OF SATELLITE COMMUNICATIONS SYSTEMS RESULTING FROM THE MOVEMENT OF THE SATELLITE RELAY STATION

§ 4.1. Non-Perturbed Motion of a Satellite in a Central Field of Gravity

In creating satellite communications systems, proper selection /95
of the orbit of the satellites, in order to assure visibility of the satellite over the entire territory where the points in the communications systems are located for a fixed time, is very important.

In this chapter, we discuss the laws of motion of satellites, basic definitions and concepts related to the motion of satellites, evolution of elements of orbits, problems of calculation of the ground tracks of movement of satellites, determination of the positions of areas of visibility, duration of mutual visibility, as well as the basic peculiarities of the transmission of signals via a moving relay satellite and other problems. In conclusions, we present a comparative analysis and basic recommendations for the selection of possible types of communications satellite orbits.

The trajectory of motion of a satellite in space (satellite orbit) is determined by many factors, the most important of which is the attraction of the satellite by the Earth. Other factors -- deceleration of the satellite in the atmosphere of the Earth, the influence of the moon, sun and planets, light pressure, the Earth's magnetic field, etc. -- also have some influence on the trajectory of the satellite. However, this influence is generally slight, and the results of this influence are the so-called perturbations in the trajectory of the satellite, i.e., deviations of the true trajectory from the ideal trajectory, calculated on the assumption that the motion of the satellite occurs under the influence of a single force -- the force of attraction toward the Earth. It should be noted here that the description of the motion of the satellite in the Earth's gravitational field, even if secondary perturbations are excluded, is a rather complex task, since the Earth is a body of complex shape with uneven mass distribution.

In particular, we can note concerning the characteristics of /96
asphericity of the Earth, that the Earth is slightly compressed along its axis of rotation, so that its equatorial radius is 6,378,245 m, while the polar radius is 6,356,864 m and, furthermore, the equatorial cross section of the Earth is reminiscent of an ellipse with slight eccentricity.

The asphericity of the Earth and the unevenness of the distribution of its mass also causes perturbations in the trajectory of satellites, but these factors can also be ignored in the first approximation, i.e., we can consider that the motion of the satellite occurs in the field of gravity of a spherical Earth with spherically symmetrical distribution of density.

The field of gravity under these conditions is called central, since it is equivalent to a field created by a material point located at the center of the Earth with mass equal to the mass of the Earth.

Let us study the unperturbed motion of a satellite in the central field of gravity of the Earth.

We select a system of rectangular coordinates with its center at the center of the Earth and axes X, Y and Z, motionless in relation to distant stars (Figure 4.1a). Suppose at a given moment in time, spacecraft S is located at a point with coordinates x_0, y_0 and z_0 at distance r_0 from the coordinate origin, and is moving with velocity V_0 in an arbitrary direction in space. If at this moment in time the motors of the spacecraft are turned off, all of its subsequent motion will be determined by the distance to the center of gravity, the magnitude and direction[1] of the velocity vector at the moment the motors were turned off and the attraction of the Earth.

According to the rules of classical mechanics, without gravity the spacecraft, after the motors are turned off, would perform even, straight motion at velocity V_0 in the direction toward which the velocity vector was directed at the moment when the motors were turned off.

When gravity is present, the spacecraft will perform a complex motion consisting in that, while continuing its linear and even motion, at the same time it falls in the direction of the center of gravity, in this case toward the coordinate origin. The force of gravity in this case gives the spacecraft a definite velocity in the direction toward the center of gravity which, when added to the vector of the initial moment, forces the spacecraft to move over a certain trajectory, the plane of which passes through the velocity vector V_0 and the coordinate origin.

If the trajectory is closed, the spacecraft will become an artificial satellite of the Earth and will continue to move on this trajectory endlessly, following the laws of motion of satellites. /97

[1]The direction is generally defined by θ_0, the angle of inclination of the velocity vector to a plane perpendicular to the radius vector of the satellite, as is shown in Figure 4.1b.

The force with which the Earth (of mass M) attracts the satellite (of mass m), separated from it by distance r, is

$$F = -k^2 \frac{Mm}{r^2}, \qquad (4.1)$$

which means that the terrestrial force of gravity gives the satellite an acceleration of

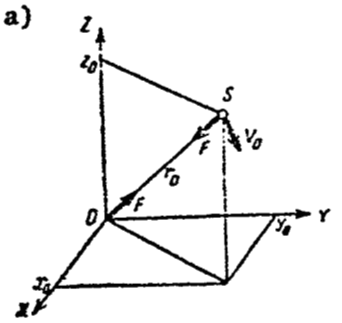
$$-\frac{k^2 M}{r^2}.$$

On the other hand, the satellite also attracts the Earth with the same force, but imparts to it the acceleration

$$-\frac{k^2 m}{r^2},$$

i.e., an acceleration as many times less as the mass of the Earth is greater than the mass of the satellite. Thus, we do not have simple rotation of the satellite about the Earth, but a more complex motion in which the Earth and satellite both rotate around a certain point near the center of the Earth. Due to the slight mass of the satellite in comparison to the mass of the Earth, the motion of the Earth can be ignored, and it can be considered that the satellite experiences both accelerations.

a) b)



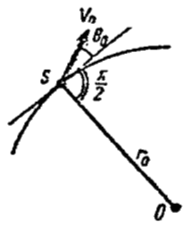


Figure 4.1. Motion of a Satellite in a Central Field of Gravity

Thus, in studying the motion of a satellite in this system of coordinates, the origin of which corresponds to the center of the Earth, we can conclude that the satellite will have an acceleration directed toward the center of the system of coordinates, equal to /98

$$-\frac{k^2(M+m)}{r^2} \approx -\frac{k^2 M 1}{r^2} = -\frac{K}{r^2},$$

where $k = 3.986 \cdot 10^5$ km^3/sec^2.

The projections of this acceleration on the axes of the coordinates give us the velocity increments along these axes:

$$\left.\begin{aligned} \frac{d^2x}{dt^2} &= -\frac{K}{r^2}\frac{x}{r} = -x\frac{K}{r^3} \\ \frac{d^2y}{dt^2} &= -y\frac{K}{r^3} \\ \frac{d^2z}{dt^2} &= -z\frac{K}{r^3} \end{aligned}\right\} \qquad (4.2)$$

The system of differential equations produced, after integration, would provide a complete solution of the problem of determination of the trajectory of motion of a satellite. However, it has been found that certain basic relationships can be determined, which immediately give us an idea of the general nature of this motion.

The most important of these relationships were empirically determined by Keppler in processing observations of the motion of the planets, and were produced mathematically strictly only much later.

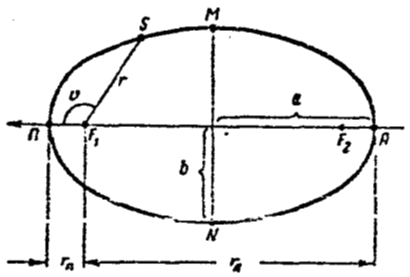


Figure 4.2. Elliptical Orbit of a Satellite (F_1, F_2, Focusses of the Ellipse, F_1 -- Center of the Earth, S -- Satellite, a -- Large Half Axis of Ellipse, b -- Small Half Axis of Ellipse)

These relationships are called Kepler's Laws and can be formulated as follows as applicable to satellites of the Earth:

Kepler's First Law.. The trajectory (orbit) of an Earth satellite lies in a nonmoving plane passing through the center of the Earth and is an ellipse, with the center of the Earth at one focus.

Kepler's Second Law. The radius vector of a satellite (straight line sector connecting the satellite and the center of the Earth) describes equal areas during equal sectors of time.

Kepler's Third Law. The ratio of the squares of periods of rotation of satellites is equal to the ratio of the cubes of their mean distances from the center of the Earth (i.e., the cubes of the the semimajor axes of the orbits).

Thus, according to Kepler's First Law, the orbit of the satellite is an ellipse, with the center of the Earth at one focus (Figure 4.2). /99

The equation for an ellipse in polar coordinates, with the center at one focus, the direction toward the closest point on the ellipse for this coordinate taken as the positive direction, is as follows:

$$r=\frac{a(1-e^2)}{1+e\cos v}=\frac{p}{1+e\cos v}, \tag{4.3}$$

where r is the radius vector of the satellite,
a is the semimajor axis of the orbit,
e is the eccentricity of the orbit ($e = F_1F_2/2a$)
v is the true anomaly (angle between radius vector and direction to the point on the elliptical orbit nearest to the center of the Earth),
$p = a(1 - e^2)$ is the focal parameter of the orbit.

Using the initial conditions, the parameters of an elliptical orbit are selected as follows:

$$p=\frac{V_0^2r_0^2\cos^2\theta_0}{K} \tag{4.4}$$

and

$$e=\frac{V_0r_0\cos\theta_0}{K}\sqrt{V_0^2-\frac{2k}{r_0}+\frac{k^2}{V_0^2r_0^2\cos^2\theta_0}}. \tag{4.5}$$

The focal parameter determines the dimensions of the orbit, the eccentricity determines its shape.

If $e < 1$, the orbit has the form of an ellipse which is more extended, the greater the eccentricity; if $e = 0$, the orbit has the form of a circle. If $e = 1$, the orbit is a parabola and, finally, where $e > 1$ -- a hyperbola.

As follows from Figure 4.2 and formula (4.3), as a satellite moves over an elliptical orbit, its distance from the center of the Earth changes continually and reaches its maximum at point A and minimum at point P. These points of the orbit are called the apogee and perigee respectively, their distance from the center

(the apogee and perigee distances) being:

$$r_A = a(1 + e)$$
$$r_p = a(1 - e). \qquad (4.6)$$

The apogee and perigee of an orbit are the ends of the major axis of the ellipse, and the line on which they are located is called the line of apsides in celestial mechanics.

The differential form of Kepler's Second Law is as follows: /100

$$r^2\frac{dv}{dt} = C, \qquad (4.7)$$

where C is a constant which depends on the initial condition (or the form an dimensions of the orbit).

In order to determine the constant, let us study two subsequent positions of the satellite, differing in time by Δt (Figure 4.3).

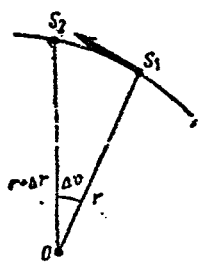


Figure 4.3. Explanation of Kepler's Second Law

The area of the sector described by the radius vector of the satellite in time Δt is

$$\Delta s \approx \frac{1}{2}(r + \Delta r) r \sin \Delta v \approx \frac{1}{2} r^2 \Delta v. \qquad (4.8)$$

Dividing both parts of the equation by Δt and turning Δt to 0, we produce

$$\lim_{\Delta t = 0} \frac{\Delta s}{\Delta t} = \frac{ds}{dt} = \frac{1}{2} r^2 \frac{dv}{dt} = \frac{C}{2}. \qquad (4.9)$$

The quantity ds/dt is called the sectoral velocity and, therefore, constant C is equal to twice the sectoral velocity.

Constant C is expressed through the initial conditions as follows:

$$C = r_0 V_0 \cos \theta_0. \qquad (4.10)$$

Kepler's Second Law allows us to judge the angular and linear velocities of motion of the satellite through its orbit. As follows from formula (4.7), the angular velocity of motion of the satellite relative to the center of the Earth at each moment in time is inversely proportional to the square of the radius vector. As concerns the linear velocity, its dependence on the position of the satellite in its orbit is somewhat different.

The velocity of motion of the satellite over its orbit can be defined as

$$V = \sqrt{V_{rad}^2 + V_{norm}^2}, \tag{4.11}$$

where V_{rad} and V_{norm} are the components of resolution vector V into two mutually perpendicular vectors -- that parallel to the radius vector and that perpendicular to it.

Since, according to Kepler's First Law,

$$V_{rad} = \frac{dr}{dt} = \frac{C}{r}\left(\frac{e\sin v}{1+e\cos v}\right),$$

and according to the Second Law /101

$$V_{norm} = C/r,$$

then

$$V^2 = \frac{K}{p}[1+2e\cos v+e^2], \tag{4.12}$$

or still simpler

$$V^2 = K\left[\frac{2}{r}-\frac{1}{a}\right]. \tag{4.13}$$

Comparing the velocities of the satellite at the apogee and perigee using these formulas, we find

$$\frac{V_A}{V_P} = \frac{1 - e}{1 + e} = \frac{r_P}{r_A}. \tag{4.14}$$

Kepler's Third Law can be produced directly from the Second Law.

As follows from formulas (4.4) and (4.10), the sectoral velocity is equal to $1/2\sqrt{pK}$, while the area encompassed by the entire elliptical orbit amounts to πab, where a and b are the semimajor and semiminor axes of the ellipse respectively.

Consequently, the time of a full rotation of the satellite through its orbit (rotation period) will be

$$T = \frac{2\pi ab}{\sqrt{pK}}. \tag{4.15}$$

Since $b = a\sqrt{1 - e^2}$ and $p = a(1 - e^2)$, we produce

$$T = 2\pi \sqrt{\frac{a^3}{K}}, \tag{4.16}$$

or

$$\frac{T^2}{a^3} = \frac{4\pi^2}{K}. \tag{4.17}$$

As follows from this formula, as the semimajor axis of the orbit increases, the period of rotation of the satellite also increases and where a = 42,200 km, it reaches 24 hours. An equatorial satellite in a circular orbit with these parameters, if launched in the direction of rotation of the Earth, will move around the Earth at an angular velocity equal to the angular velocity of rotation of the earth itself. In this case, the satellite will be stationary relative to the surface of the Earth, which is very convenient from the point of view of organization of communications. The area of visibility of such a satellite is quite great (about one-third of the entire surface of the Earth) and is constant; furthermore, the immobility of the satellite simplifies tracking and makes it easy to transmit signals through the satellite. This type of satellite, called stationary, is used in many existing and planned satellite communications systems. /102

Thus, we have studied Kepler's three laws in detail, describing the nature of the motion of a satellite as a function of its position in its orbit, and have determined the relationship of the basic parameters of the orbit to the initial conditions. As we can see from formulas (4.3), (4.7) and (4.17), the mathematical expression of Kepler's laws, none of these laws gives us the direct connection between orbital parameters and time. This connection, namely the dependence of the true anomaly on time, can be defined by integration of the expressions for the first and second laws.

The integration produces

$$\frac{\pi t}{T} = \operatorname{arctg}\left(\frac{\operatorname{tg}\frac{v}{2}}{\sqrt{\frac{1+e}{1-e}}}\right) - \frac{e\sqrt{\frac{1+e}{1-e}}\operatorname{tg}\frac{v}{2}}{\frac{1+e}{1-e}+\operatorname{tg}^2\frac{v}{2}} \tag{4.18}$$

i.e., the inverse of the function t(v), which makes this formula difficult to use. In practice, the expansion of v(t) into a series with respect to Bessel functions or other artificial methods are frequently used.

Thus, considering everything we have said, we can calculate the nonperturbed motion of an Earth satellite relative to the plane of its orbit.

The motion of a satellite relative to an observer on the ground is more complex, since the observations of the satellite are conducted from the surface of the Earth, which is rotating in space relative to the plane of the orbit.

§ 4.2. Elements of the Orbit of a Satellite

As we know from celestial mechanics [4.1], the motion of a material point in a fixed field of forces is described by a system of ordinary sixth order differential equations. It follows from this that the orbit of a satellite can be fully defined by six independent constants, i.e., six independent parameters called the elements of the orbit allow us to find the coordinates of the satellite and the components of the velocity vector of the satellite at any point on its orbit at any moment in time.

In order to determine the basic elements of the orbit, let us select a system of coordinates (Figure 4.4) formed by the plane of the equator of the Earth and the axis of rotation of the Earth (positive direction from the center of the Earth to the north pole). Axis X in this case will be directed from the center of

the Earth to the point of intersection of the Greenwich meridian with the equator, axis Y -- to the point with coordinates 0° s. lat., 90° e. long.

Let us now define the six basic elements of the orbit. /103

The first two -- the magnitude of the semimajor axis and the eccentricity of the orbit e -- have already been mentioned. These elements define the dimensions and shape of the orbit, as well as the motion of the satellite in the plane of the orbit.

Another two -- the longitude of the ascending node Ω and the inclination of the orbit i -- determine the position of the plane of the orbit in space.

The inclination of the orbit i is measured as the angle between the Z axis and line N, passing through the coordinate origin perpendicular to the plane of the orbit, with its positive direction in that direction from which the motion of the satellite is observed to be counterclockwise. The inclination may vary between 0 and 180°, with satellites moving in the direction of rotation of the Earth having inclinations of 0 to 90°.

At each moment in time, the orbit of the satellite intersects the plane of the terrestrial equator at two nodal points. Passing through these points, the satellite moves from the northern hemisphere into the southern hemisphere or from the southern hemisphere into the northern hemisphere. The first point is called the descending node, the second point -- the ascending node. The longitude of the ascending node Ω is read off as the angle at the plane of the equator from the X axis counterclockwise (observing from the positive direction of the Z axis) to a straight line connecting the coordinate origin to the ascending node (line of nodes). Since the plane of the orbit retains its position relative to the stars, while the Earth rotates relative to the stars, the longitude of the ascending node changes evenly with time within limits of 0 to 360°. As an element of the orbit, the longitude of the ascending node relates to some definite moment in time, for example the moment when the satellite passes through the perigee t_0.

Thus, t_0 is the fifth element of an orbit, which relates both the position of the plane of the orbit in space and the position of the satellite in the orbit to time.

The last, sixth, element of an orbit is the argument of the perigee or the angular distance of the perigee from the ascending node ω, which is read off as the angle in the plane of the orbit from the direction to the ascending node to the direction to the perigee. The perigee argument may be any angle between 0 and /104

360°. Satellites with their apogee in the northern hemisphere should have a perigee argument of between 180 and 360°.

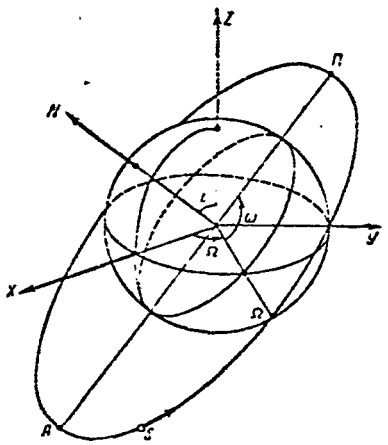


Figure 4.4. Elements of the Orbit of a Satellite

§ 4.3. Evolution of Orbital Elements

As was noted above, the position of the orbit in space remains unchanged if only a central gravitational force acts on the satellite. Actually, the satellite is subjected to the influence of various perturbing factors: noncentrality of the field of gravity of the Earth, attraction of other planets, deceleration of the terrestrial atmosphere, the pressure of sunlight, etc., which distort the initial position of the orbit as time passes, change its elements and, in turn, lead to a change in the conditions of satellite communications (displacement of the area serviced, changes in communications session scheduling, etc.).

Within the sphere of influence of the Earth (but outside the dense layers of the terrestrial atmosphere), perturbing forces are slight in comparison to the basic force of gravity. However, they may have a noticeable influence on the movement of a satellite in orbit. The most significant perturbing forces are the so-called secular perturbations of an orbit, continually increasing with flight time of the satellite and (even when the perturbing forces are comparatively slight), significantly changing the entire nature of motion over a long interval of time.

In this section, we will describe the secular perturbations of an orbit, caused by the asphericity of the Earth and the attraction of the moon and the sun.

The presence of the terrestrial atmosphere can cause significant distortions of the orbit of a satellite at orbital altitudes of less than 500 km, if the satellite is light and has a large cross section. In the same manner, the pressure of sunlight can cause a significant distortion of the orbit only if the satellite is light and has a large cross section.

Let us study the perturbations in orbital elements caused by the flattening of the Earth. The study of the equations of motion of a satellite in the actual field of gravity of the Earth

allows us to find expressions (in the first approximation) for the changes in elements of an elliptical orbit in one rotation of the satellite [4.2]. These expressions are as follows

$$\left.\begin{aligned} \delta p = \delta e = \delta i = 0 \\ \delta\omega = \frac{\pi}{p^2}\frac{\alpha}{K}(5\cos^2 i - 1) \\ \delta\Omega = -\frac{2\pi}{p^2}\frac{\alpha}{K}\cos i \end{aligned}\right\} \qquad (4.19)$$

where $\alpha = 0.266360022 \cdot 10^{11}$ km^5/sec^2 is a quantity characterizing the flattening of the Earth

Analyzing the formula for determination of the secular change in perigee $\delta\omega$, we note that when the inclination is /105

$$i_1 = \arccos\sqrt{\frac{1}{5}} \approx 63°26',$$

$$i_2 = \pi - i_1 \approx 116°34',$$

the value of $\delta\omega = 0$, i.e., the rotation of the line of apsides (semimajor axis of the elliptical orbit) is zero. Where $i < i_1$ or $i > i_2$, we produce $\delta\omega > 0$, i.e., the line of apsides rotates in the direction of motion of the satellite. Where $i_1 < i < i_2$, on the other hand, $\delta\omega < 0$, i.e., the perigee is displaced in the direction opposite to the direction of motion of the satellite. The maximum rate of rotation of the line of apsides corresponds to equatorial orbits ($i = 0$ or π).

Going over to analysis of the formula for the secular drift of the line of nodes (the so-called regression of the line of nodes), we note that the line of nodes is gradually displaced in the direction from east to west where $0 \leq i < (\pi/2)$ and from west to east where $(\pi/2) < i \leq \pi$. For polar orbits ($i = \pi/2$), there are no secular changes of the line of nodes. As the orbit approaches the plane of the equator, the rate of the secular change in nodes increases.

Upon transition from an elliptical orbit to a circular orbit, the focal parameter p becomes equal to the radius r of the circular orbit. The formula for determination of the secular displacement of the perigee $\delta\omega$ becomes senseless in this case, while the formula defining the displacement of the nodes is converted to the corresponding formula for circular orbits:

$$\delta\Omega = -\frac{2\pi}{r^2}\frac{a}{K}\cos i. \qquad (4.20)$$

Thus, the asphericity of the Earth causes changes in two elements of the orbit: rotation of the line of apsides (with inclinations other than $i_1 = 63°26'$ and $i_2 = 116°34'$) and rotation of the line of nodes (with inclinations other than $i = 90°$). The remaining elements remain practically unchanged, i.e.,

$$\delta p = \delta e = \delta i = 0.$$

Let us now study the perturbing influence of the attraction of the sun and moon on the motion of satellites. It has been proven that the perturbing influence of celestial bodies, with the exception of the sun and the moon, on the motion of satellites at altitudes of less than 100,000 km is negligible. For satellites flying at altitudes on the order of several thousands of kilometers or less, the perturbations of the orbit by the sun and the moon are significantly less than the perturbations resulting from the asphericity of the Earth. Beginning at an altitude of about 50,000 km, the perturbations resulting from the attraction of the sun and the moon exceed all other gravitational perturbations. The perturbations resulting from the attraction of the moon are approximately 2.2 times the perturbations resulting from the attraction of the sun. /106

Analysis of the orbital perturbations shows that the attraction of the moon and the sun causes a change in the position and altitude of the perigee (when the semimajor axis is constant, i.e., $\delta a = 0$), as well as rotation of the plane of the orbit about an axis which is a projection of the direction toward the perturbing body in the plane of the orbit.

Particular attention should be given to the evolution of the height of the perigee, since changes in orbital inclination, position of nodes and perigee argument are insignificant and involve only changes in orbital geometry, whereas a reduction of the altitude of the perigee might cause the satellite to fall back to Earth. The height of the perigee of the orbit increases if the angle formed by the projection of the vector defining the position of the perturbing body relative to the Earth on the plane of the orbit and the direction toward the perigee of the orbit lie in the first or third quadrant, and decreases if the angle lies in the second or fourth quadrant. Thus, the evolution of orbital elements of a satellite is significantly influenced by the date and time of launch of the satellite and the related values of longitude of the ascending node, which determine the initial mutual position of perturbing extraterrestrial bodies and the orbit in space.

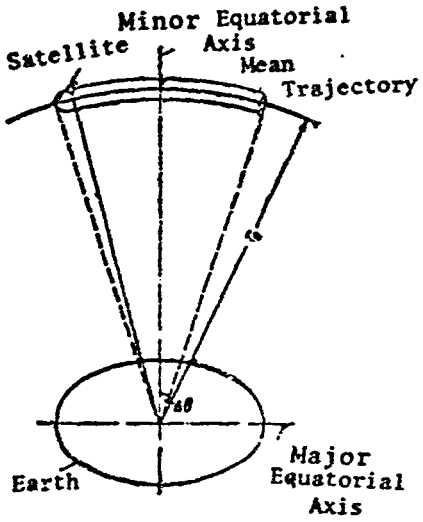


Figure 4.5. Geometric Plan of Mean Trajectory of Motion of a Stationary Satellite

The change in perigee altitude H_p depends little on its initial value and is primarily determined by the altitude of the apogee H_a. With elliptical orbits as eccentricity decreases, $\delta H_p \to 0$.

The primary influence on the evolution of the orbit of a stationary satellite is that of the equatorial compression of the Earth [4.3]. Due to the flattening of the Earth at the equator, the vector of the force of gravity is deflected from the center of the Earth in the direction of the large axis of the equatorial ellipse (Figure 4.5). The gravitational attraction can be resolved into two components: vertical and horizontal. The latter component is quite slight and has the same influence on a satellite as a slight thrust or aerodynamic drag, in that it causes harmonic oscillations of altitude and velocity of motion of a stationary satellite, causing periodic "drift" of the satellite along the equator about the minor axis of the equatorial ellipse.

The period and amplitude of oscillations are determined by /107
the difference in longitudes of the position of the minor axis of the equatorial ellipse and the initial position of the satellite (its point of insertion). For the case of ideally precise insertion of the satellite, the amplitude of oscillations will be equal to the difference in these longitudes $\Delta\theta$ and the dependence of oscillating period T on amplitude will be as shown in Figure 4 6.

The gravitational fields of the moon and sun cause long-period changes in the inclination of the orbit of the stationary satellite, with period of 200 and 600 years respectively. In one year, the change in orbital inclination is approximately 0.92°. This means that one year after insertion of a satellite at a fixed point, it will describe a "figure eight" against the sky with a meridianal length of approximately ±0.92°. The maximum correcting velocity necessary to eliminate this "drift" of the satellite due to equatorial compression of the Earth is about 2.1 m/sec per year, whereas correction of the inclination change resulting from the fields of gravity of the moon and sun

requires about 54 m/sec per year. Therefore, correction of the orbital inclination of a stationary satellite is desirable.

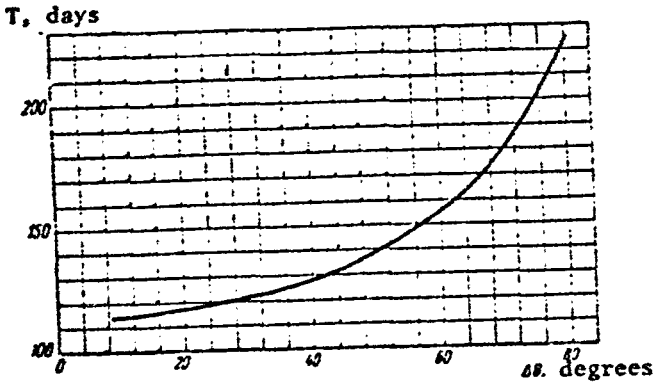


Figure 4.6. "Drift" of a Stationary Satellite as a Function of Maximum "Drift" Angle

§ 4.4. Calculation of the Track of Motion and Area of Visibility of a Satellite

In order to describe the motion of a satellite relative to an observer on the Earth, the geocentric spherical system of coordinates in which the position of the satellite at each moment in time is determined by the geographic coordinates (latitude and longitude) of the point on the surface of the Earth for which the satellite is located at the zenith (point beneath the satellite) and the radius vector of the satellite is most convenient. /108

The coordinates of the point beneath the satellite can be calculated by the following formulas

$$\phi_{1ts} = \arcsin\{\sin i \sin[\omega + v(t-t_0)]\}. \quad (4.21)$$

$$\theta_{1gs} = \Omega_0 - \frac{2\pi}{T_{sb}}(t-t_0) + \operatorname{arctg}\{\cos i \operatorname{tg}[\omega + v(t-t_0)]\}. \quad (4.22)$$

where ϕ_{1ts} is the latitude of the point beneath the satellite,
θ_{1gs} is the longitude of the point beneath the satellite,
Ω_0 is the longitude of the ascending node at moment t_0,
t is the current time,

t_0 is the moment of passage through the perigee,

T_{sb} is the duration of a stellar day.

With known satellite orbital elements, the ground track of the point beneath the satellite can be constructed, the moment when a satellite passes over a given point on this track can be determined and the zone of visibility of the satellite can be calculated for each moment in time.

If we limit the zone of visibility of the satellite to that portion of the Earth's surface for which the satellite is visible at not over a certain fixed angle β_{min}, then (see Figure 4.7), the following relationships are correct between the radius vector (r), maximum slant altitude (d_{max}), angle α, representing the angular distance of the boundary of the zone from that point beneath the satellite S' (relative to the center of the Earth) and angle γ, equal to half the maximum angular size of the zone of visibility relative to the satellite, at each moment in time:

$$\left.\begin{aligned} &\gamma+\beta_{min}-\alpha=90^\circ \\ &d_{max}^2=r^2+R_3^2-2R_3r\cos\alpha \\ &d_{max}=\frac{R_3\sin\alpha}{\sin\gamma}=\frac{r\sin\alpha}{\cos\beta_{min}} \end{aligned}\right\} \qquad (4.23)$$

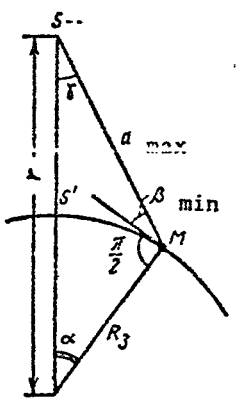


Figure 4.7. Definition of Zone of Visibility

In the problem of determining the zone of visibility, the moment in time is generally fixed (i.e., r and the coordinates of the point beneath the satellite φ_{1ts} and θ_{1gs} are known), as well as β_{min}.

The coordinates of the boundary of the zone of visibility in this case can be calculated in the following sequence:

$$\sin\gamma=\frac{R_3\cos\beta_{min}}{r},$$

from which

$$90^\circ-\beta_{min}-\arcsin\frac{R_3\cos\beta_{min}}{r},$$

$$\alpha = 90^\circ - \beta_{min} - \arcsin \frac{R_3 \cos \beta_{min}}{r}$$

the coordinates of the boundary (ϕ_{1tm} and θ_{1gm}) are found from the equation

$$\cos \alpha = \sin \phi_{1ts} \sin \phi_{1tm} + \cos \phi_{1ts} \cos (\theta_{1gs} - \theta 1gm). \quad (4.24)$$ /109

Fixing various ϕ_{1tm} within the limits of $\phi_{1ts} \pm \alpha$, we find the corresponding θ_{1nm} and use these data to construct the zone of visibility.

If the pole is included in the zone of visibility, $\phi_{1ts} + \alpha$ will be greater than 90° and, in calculating the upper limit for ϕ_{1tm}, we take it as 180° - (ϕ_{1ts} + α).

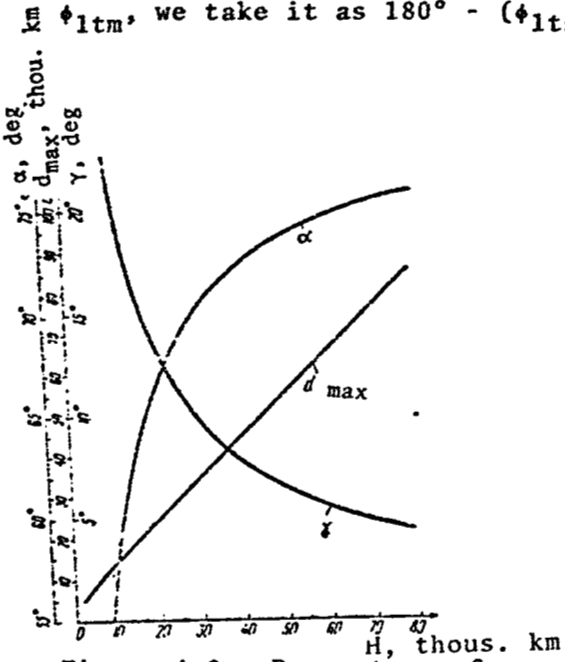


Figure 4.8. Parameters of Zone of Visibility as Functions of Altitude (β_{min} = 10°)

Figure 4.8 shows the dependence of the basic parameters of the zone of visibility on altitude of a satellite over the surface of the Earth. The altitudes were taken within limits of 0 to 80,000 km, i.e., up to the maximum altitude which a satellite can have on an elliptical or circular orbit with a period of rotation of up to 24 hours. We can see from the figure that the angular dimensions of the zone of visibility increase sharply with increasing orbital altitude of the satellite, i.e., considering that satellites at higher altitudes usually move more slowly, the duration of communications sessions between any fixed points within the zone of visibility of the satellite increases still more sharply. However, at very high altitudes various difficulties arise, related to the great time of propagation of a signal between the Earth and the satellite and the great attenuation of signals, which limits the altitude of the communications satellites to 20-40 thousand km.

As the satellite moves through its orbit, both the position and the dimensions of the zone of visibility generally change. In this case in estimating the length of a communications session for fixed communications points on the map, we can draw in the trajectory of the point beneath the satellite and several successive areas of visibility. In this manner, we can easily find the maximum duration of a communications session for fixed points and determine which of the orbital elements must be changed and in what direction in order to provide the required duration of a session.

Figure 4.9 presents the trajectories of the points beneath the satellite for satellites in elliptical orbits with rotation periods of 8, 12 and 24 hours. The inclination of the orbit is selected optimal from the point of view of lack of perturbations ($i = 63.4°$), the perigee altitude is taken as 500 km and the perigee argument as 270°, so that the satellite is in the northern hemisphere most of the time and moves as slowly as possible at high altitudes, providing the maximum attainable duration of a communications session for an orbit with the assigned period and inclination. The trajectories are drawn on the map in cylindrical projection and relate to one rotation of the satellite. For subsequent rotations, the picture is repeated with a shift of $15° \cdot T_{rev}$(hr) in longitude. The moment when the satellite passes through its perigee is taken as the zero point on the time scale. The trajectories are matched so that the points corresponding to the moment of passage through the apogee are on the zero (Greenwich) meridian. If the longitude of the apogee is other than zero, the entire picture can be simply shifted in longitude, due to the projection used. /110

As an illustration of the change in size and position of zone of visibility with time, Figure 4.10 shows the trajectory of the point beneath the satellite and zones of visibility for various moments in time for the Molniya-1 satellite with its 12-hour elliptical orbit with $H_p = 500$ km, $\omega = 270°$ and $i = 63.4°$.

As follows from the figure, the satellite, as it moves through this orbit is located in the southern hemisphere only about one hour per revolution (12 hours), at which time it moves quite rapidly at low altitude, i.e., is very inconvenient for communications. As the satellite moves near its apogee, due to its low velocity and great altitude, it is visible simultaneously from many points over a large area of the northern hemisphere for a long time (up to 9-10 hr).

The shape of the trajectory indicates how slowly a satellite moves at its apogee. In this case, the angular velocity of displacement at the apogee is less than the angular velocity of the daily rotation of the Earth and the satellite "lags" behind the

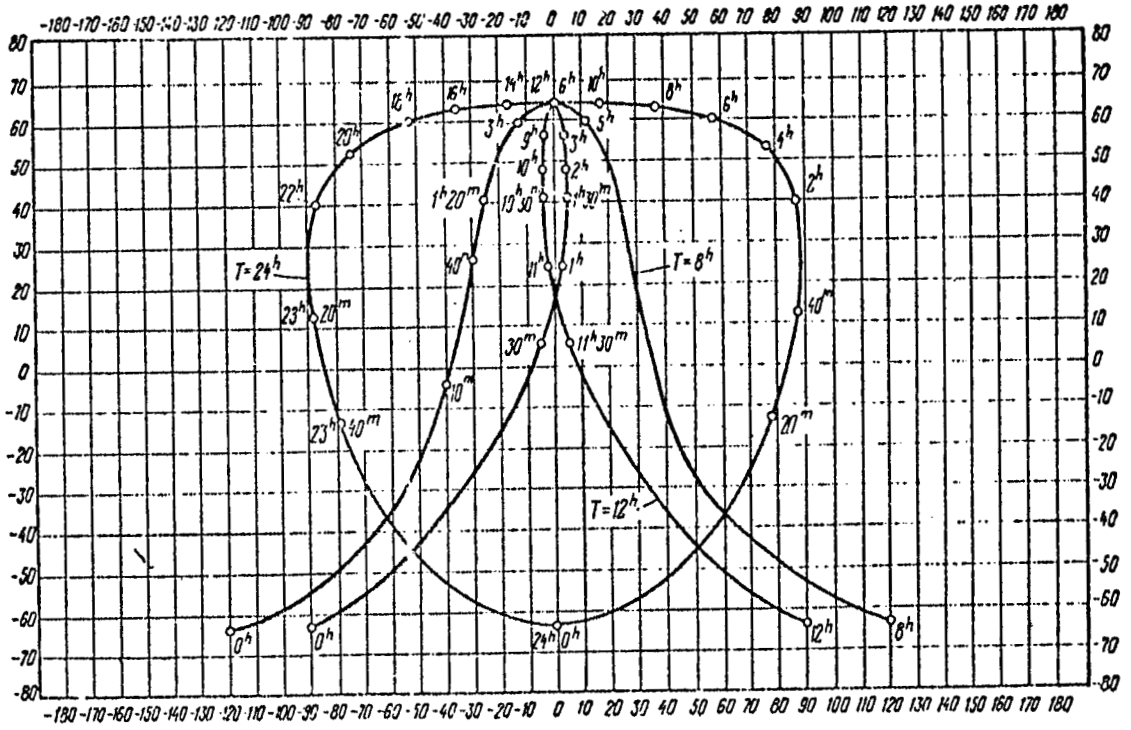


Figure 4.9. Trajectory of Point Beneath Satellite for Satellites in Elliptical Orbits with Rotation Periods of 8, 12 and 24 Hours (H_p = 500 km, ω = 270°, i = 63.4°)

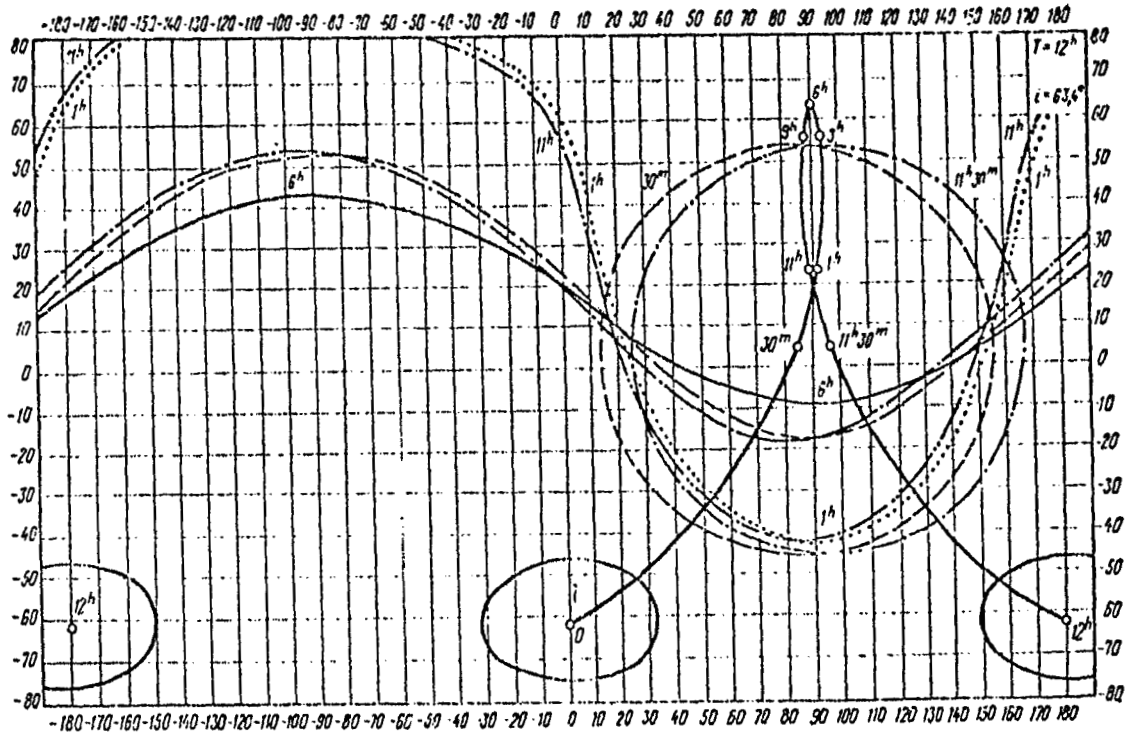


Figure 4.10. Trajectory of Point Beneath Satellite and Zones of Visibility for a Satellite in an Elliptical Orbit with Rotation Period 12 hours (H_p = 500 km, ω = 270°, i = 63.4°)

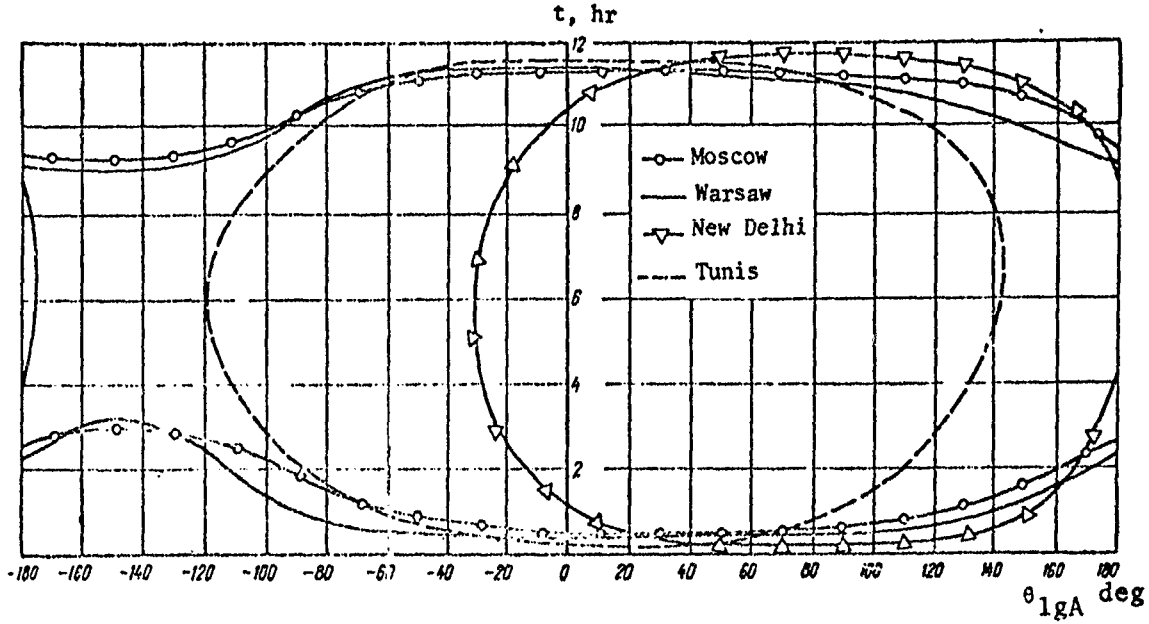


Figure 4.11. Moments of Beginning and Ending of Visibility of a Satellite in an Elliptical Orbit with a Rotation Time of 12 hours (H_p = 500 km, ω = 270°, i = 63.4°) as Functions of Apogee Longitude for Moscow, Warsaw, New Delhi and Tunis

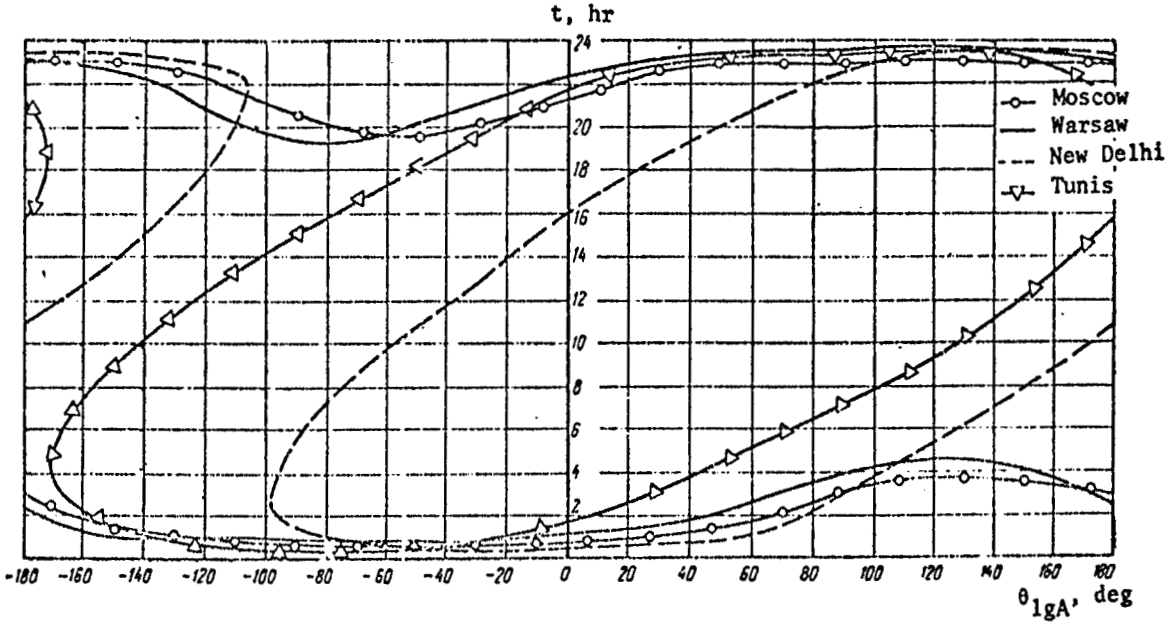


Figure 4.12. Moments of Beginning and Ending of Visibility of a Satellite in an Elliptical Orbit with a Period of Rotation of 24 hours (H_p = 500 km, ω = 270°, i = 63.4°) as a Function of Longitude of Apogee for Moscow, Warsaw, New Delhi and Tunis

/114

Earth, moving relative to an observer in the direction opposite to the direction of rotation of the Earth.

Due to the great communications session duration provided by a satellite in an orbit with these parameters for a large territory, and also due to the convenience of launching of these satellites from the territory of the USSR, these satellites, as we know, are widely used in the USSR for transmission of television programs in the Orbita system and for experiments involving transmission of television and telephone conversations over extremely long distance lines (Moscow-Vladivostok, Moscow-Paris).

As follows from Figure 4.10, the duration of a communications session between two or more points depends on their geographic position and the location of the orbit of the satellite relative to these points. Communications between all points simultaneously, naturally, can be begun as soon as a satellite appears over the horizon for the last of the points and must be completed by the time the satellite leaves the zone of visibility for any one of the points. /115

Figures 4.11 and 4.12 show the moments of beginning and ending of visibility of a satellite in 12 and 24 hour elliptical orbits with the same parameters as on Figure 4.9 as functions of longitude of apogee (θ_{lga}) for Moscow, Warsaw, New Delhi and Tunis. These figures allow us to select the optimal values of apogee longitude to achieve the maximum time of simultaneous communications for two or more points.

§ 4.5. Peculiarities of Transmission of Signals through Moving Relay Satellites

In a communications system using satellites, the signal received by a subscriber generally undergoes the following changes:

-- Doppler frequency shift, caused by the relative movement of the satellite and ground stations;

-- delay of the signal as it is transmitted from Earth to satellite to Earth, the duration of which is always greater than for stations on the surface of the Earth;

-- variations in delay and Doppler shift at the moment when a signal is switched from one satellite to another, caused by differences in the Earth-satellite-Earth track at the moment of switching and different relative velocities of satellites.

4.5.1. Doppler Frequency Shift

The Doppler frequency shift over the sector from Earth to satellite and from satellite to Earth depends on the operating

frequency over the sector and the velocity of the satellite relative to the ground station. If for example, the satellite is moving away from the ground station at V_{dop}, the electromagnetic oscillation radiated by the ground transmitter at frequency f will be received on board the satellite as an oscillation at frequency

$$f_{dop} = f(1 + (V_{dop}/c)), \qquad (4.25)$$

where c is the propagation velocity of the electromagnetic oscillations.

When modulated oscillations are transmitted, the frequency of each spectral component of the signal changes by $(1 + (V_{dop}/c))$ times, i.e., the components in the upper sideband receive a greater absolute frequency increment than those in the lower sideband.

For example, in the simplest case, if a signal consists of a carrier of frequency f_0 and two symmetrically placed sidebands $f_0 - F$ (lower sideband) and $f_0 + F$ (upper sideband), the Doppler effect will cause the frequencies to take on the following values: /116

$$f_0\left(1 + \frac{V_{dop}}{c}\right) = f_0 + f_0\frac{V_{dop}}{c} = f_0 + \Delta f_0$$

$$(f_0 - F)\left(1 + \frac{V_{dop}}{c}\right) = (f_0 + \Delta f_0) - F\left(1 + \frac{V_{dop}}{c}\right)$$

$$(f_0 + F)\left(1 + \frac{V_{dop}}{c}\right) = (f_0 + \Delta f_0) + F\left(1 + \frac{V_{dop}}{c}\right). \qquad (4.26)$$

The results produced can be represented as two phenomena: a general shift of the entire spectrum by Δf_0 and a change in the scale of the spectrum by $(1 + (V_{dop}/c))$ times, i.e., deformation of the spectrum.

The shift of the entire spectrum or the Doppler shift of the carrier frequency must be considered in calculating the transmission band of satellite and ground station equipment; compensation of this shift is not difficult.

The deformation of the spectrum, as we can see from formula (4.26), causes each component of the spectrum of the modulating oscillation to change its frequency by $(1 + (V_{dop}/c))$ times after

transmission through the satellite and demodulation, and compensation for this change, if necessary, is quite difficult.

The value of the relative Doppler frequency shift and degree of deformation of the spectrum are determined to a great extent by the parameters of the orbit of the satellite and the placement of the ground stations; depending on the situation, the Dopper frequency shift on the two sectors of the transmission path may be added or subtracted. The maximum Doppler shift upon transmission of signals between several ground stations will be observed for pairs of stations located near each other for which the relative velocity of the satellite is greatest.

Calculation of the Doppler shift with fixed parameters of the orbit and position of the ground station can be performed by differentiation of the expressions for ground station-satellite distance with respect to time. In the general case, these are rather cumbersome calculations, which need not be presented here.

Figure 4.13 shows the results of calculation of the maximum and minimum relative Doppler shifts as functions of time (rotation of the Earth ignored) for a satellite in a 12 hour elliptical orbit. This is the orbit used by Molniya satellites, is greatly elongated, and the velocity of the satellite at the perigee is significantly greater than its velocity at apogee. However, since the satellites are switched on at altitudes of 15-20 thousand km, the area of greatest values of Doppler shift is not utilized.

The maximum Doppler frequency shift encountered in transmission over one of the sectors of the link with a satellite in a circular orbit can be determined from the following relationship: /117

$$\Delta f_0 \approx \pm 1{,}5 \cdot 10^{-6} f_0 s, \qquad (4.27)$$

where s is the number of rotations of the satellite around the Earth per day.

Table 4.1 shows the value of s for several circular equatorial orbits of satellites.

Thus, if we take an altitude of 5,000 km as the minimum altitude of a circular orbit used in a communications system, the maximum relative Doppler shift and spectral deformation will be about $5 \cdot 10^{-6}$. Actually, even higher circular orbits are used in communications systems and, consequently, the influence of the Doppler effect is still less.

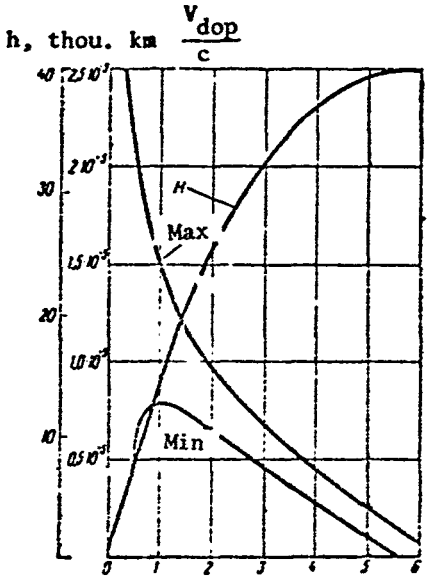


Figure 4.13. Change in Maximum and Minimum Relative Doppler Shift, as well as Altitude of Satellite, as it Moves over a Molniya-type Orbit

However, when narrow band signals are transmitted, one must be sure that each station is allocated a band of frequencies considering possible Doppler shift in both directions, and that it is possible to filter out the two neighboring signals with the decreased difference between them at the two maximum Doppler shifts. This requires that the guard band between HF channels be broadened and that the total band width occupied by a communications system be increased when there are many such channels.

For a numerical estimate of these effects, Table 4.2 presents the values of maximum Doppler shift of a signal at 6 GHz for satellites in circular equatorial orbits with various altitudes (satellites launched in the direction of rotation of the Earth).

It follows from the table that if the system is designed for transmission of 20 HF channels, the band of frequencies occupied by the system considering the Doppler effect must be expanded in the worst case by 1 MHz. As concerns spectral deformation, it is most strongly felt at the upper frequencies of the frequency band of the modulating signal or in the upper channels of multichannel telephone systems with frequency-division multiplexing. /118

TABLE 4.1

Number of Rotations of Satellite around Earth per Day, s	0	1	2	3	4	5	6	7
Altitude of Circular Equatorial Orbit, km	35600	20240	13940	10390	8080	6420	5170	4190
Period of Rotation of Satellite, hr	24	12	8	6	4.8	4	3.4	3

TABLE 4.2

Period of Rotation, hr	6	8	12	24
Altitude, thous. km	11	14	20	36
Maximum Doppler Shift ($f = 6$ GHz, $\beta_{min} = 5°$)	27 KHz	18 KHz	9 KHz	0.12[1] KHz

[1]This value is calculated on the assumption that perturbations of the satellite cause it to deviate from its fixed position by not more than 60 km. In this case, the maximum velocity of the satellite relative to the ground station will be not over 5.8 m/sec.

According to recommendation G.225 of the CCITT, the difference in audio frequencies at the receiving and transmitting ends should not exceed 2 Hz. This requirement places a limitation either on the width of the group frequency band transmitted, or on the velocity of motion of the satellite relative to ground stations. If it is impossible or undesirable to follow these limitations, one of two known methods for compensation for Doppler shift should be used.

The first method involves the use of a device with a variable, controlled delay of the group HF signal. This method achieves full compensation for the influence of the Doppler effect, since the action of the device with variable delay fully imitates the process of development of Doppler shift, and can be directed in the opposite direction. The delay can be controlled either according to a program or by a special pilot signal transmitted by the ground station through the satellite and received at the same station.

The second method of compensation involves control of the /119
frequencies of the multiplexing apparatus. This method provides precise placement of the channels in the group band of frequencies, but does not eliminate the deformation of the spectrum of the message itself.

If the recommendation G.225 of CCITT is observed, there will be no difficulties in the transmission of telegraphy and data through telephone channels. When television signals are transmitted, the change in frame frequency caused by the Doppler effect is very slight and can be ignored for black and white television for ordinary home television receivers.

It was originally assumed that compensation of the Doppler effect might be required for transmission of color television signals by satellite. However, successful experimental

transmissions have shown that standard color television receivers operate satisfactorily at the Doppler shift levels expected in actual satellite communication systems.

4.5.2. Signal Delay

In satellite communications systems signals transmitted are received by ground stations with a significant delay, caused by the long propagation times over the path from the Earth to the satellite and back to the Earth. This delay increases still more when the signal passes through several such lines (steps) in sequence.

An increase in satellite altitude also increases the duration of the delay and for a stationary satellite, located at an altitude of about 36,000 km, this delay reaches 300 msec. Since it is improbable that satellites will be used at still greater altitudes, we can consider that 300 msec is the maximum delay, and is one of the shortcomings of stationary satellites. However, for a stationary satellite this delay is constant for any given combination of ground stations, whereas when satellites in other orbits are used the delay changes with the passage of time, causing certain inconveniences.

For a one-hop line, the minimum possible delay time t_{min} occurs when two ground stations located close together communicate through a satellite located near the zenith for both stations, where

$$t_{min} = \frac{2H}{c}, \tag{4.28}$$

where H is the altitude of the satellite over the Earth,
c is the speed of light.

The maximum possible delay will be observed when the satellite is on the horizon for both stations ($\beta_{min} \to 0$). The maximum /120 delay can be found from the following relationship:

$$t_{max} = \frac{2H}{c}\sqrt{1 + 2(R_g/H)}. \tag{4.29}$$

In practice, the minimum delay is slightly over t_{min} as determined by formula (4.28), since the stations will be separated by some distance, while the maximum delay will be somewhat less than t_{max}, as defined by formula (4.29), since it is not likely that both stations will see the satellite simultaneously at a very

slight angle, and in practice this angle is limited to 5-10°. However, the expressions can give us a good idea of the possible limits of change of delay.

In calculating the delay time, we must keep in mind that some delay of the signal also occurs as it passes over the surface communications lines. Usually this delay is not great, but when low-altitude satellites are used with extensive connecting lines, the delay on the surface lines may represent a significant portion of the total delay for ground stations.

In recommendation G114 of ITTCC, it is suggested that signal travel time over surface lines by calculated by the formula

$$t_{con} = 12 + 0.004\,\ell\,(\text{msec}), \qquad (4.30)$$

where ℓ is the length of the connecting lines, km.

In estimating the delay time, it must be considered that it is not probable that long connecting lines will be used simultaneously at both ground stations, so that a time of 30 msec can be used as the average delay time of the signal at both ends, with a maximum of 50 msec, a minimum of 10 msec. Considering these facts, based on formulas (4.28) and (4.29), we have calculated the minimum, maximum and mean delays for a one-hop line (in one direction) for satellites in various orbits, and the results are presented in Table 4.3.

TABLE 4.3

Communications Line Sectors	Signal Delay, msec								
	H = 11 thou km			H = 14 thou km			H = 36 thou km		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
E-S-E	74	109	92	92	128	110	240	280	260
Connecting lines	10	50	30	10	50	30	10	50	30
Total delay	84	159	122	102	178	140	250	330	290

For a multihop connection, the mean delay will increase approximately in proportion to the number of hops, the maximum delay -- somewhat more slowly. /121

In communications systems using moving satellites, the delay of a signal in any specific line will change slowly with time as the satellite moves. In practice in systems using satellites in circular orbits with orbital periods of over 6 hours, the delay generally changes within limits of 20 msec. A variable-delay device can be used to compensate for changes in delay or, at least,

decrease these changes. The delay introduced by this device to the signal path must be continuously adjusted according to a program or a pilot signal so that the signal always has a constant delay, near the maximum value.

The change in delay with time and variation in delay for all stations strongly influence the operation of a multistation access system with time division multiplexing. With multistation access with time division multiplexing, time is divided into intervals, and a time sector is set aside for transmission from each station within each interval (see Chapter 8). In these systems, very high requirements are placed on the synchronism of operation of all stations. A change in delay time of the signals of each station by any significant amount hinders synchronization of the operation of all stations in the network and, even when synchronization is achieved, certain difficulties arise in the reception of signals if their time of arrival is not fully defined. If the delay time of a signal over the sector from station A to the satellite is t_A, from station B to the satellite is t_B, the difference in these times $\Delta t = t_A - t_B$ defines the position in time of the signals from stations A and B as received at the satellite.

When multistation access systems with time division multiplexing are used with a stationary satellite, the signal channel of each station must include a constant delay device in order that Δt will be the same for all stations. In the case of moving satellites, the delay introduced must be variable or, if delay is not regulated, the time intervals between two signals must be increased.

Table 4.4 shows the maximum values of Δt for satellites in circular orbits at various altitudes.

The absolute value of delay for high altitude satellites becomes noticeable in telephone conversations and causes certain difficulties for the subscribers.

Several experimental studies have been performed in order to determine just how noticeable delay is in telephone conversations. Based on results of these studies, the ITTCC has proposed the following preliminary recommendation:

-- delays in one direction of 0 to 150 msec are unconditionally acceptable; /122

-- delays of between 150 and 400 msec in one direction are temporarily acceptable;

-- delays of over 400 msec in one direction are temporarily unacceptable. Connections with delays of this length should be avoided in planning.

TABLE 4.4

Orbital Period, hr	6	8	12	24
Altitude, thous. km	11	14	20	36
Δt, msec	30.5	32.0	33.7	0.124[1]

[1]See note to Table 4.2.

Studies of the influence of delay on telephone conversations have also shown that with great delays, one source of discomfort for subscribers is the echo signals which are heard, caused by imperfect decoupling of the reception and transmission channels at both ends of the line. When delays are great, the echo signals lose their correlation with the initial signal and sound like a second conversation on the line.

Many types of echo supressors which operate rather effectively have been developed for the echo signal problem. Their operating principle consists in automatic breaking of the reception circuit for the subscriber speaking.

4.5.3. Interruption of Signal Continuity at the Moment of Switching

In systems with moving satellites, communications are conducted through several satellites, which successively pass over the territory where the ground stations are located. When one satellite moves beneath the horizon and the next rises above the horizon, there is a time during which all stations must simultaneously re-aim their antennas from the first satellite to the second. This maneuver, when many stations are involved, may take a comparatively long period of time, during which signals cannot be transmitted. To avoid interruptions to communications, each station must have two antennas, so that when the moment of switching arrives the second antenna can already be pointed at the new satellite to allow communications with all other stations, and for a certain period of time two identical stations operate simultaneously, transmitting the same signals. /123

After communications are established through these second antennas, all stations are switched and the first antennas then become the second antennas, awaiting the next switching operation. Effective operation of this system requires that both satellites (the rising and setting satellites) be in the area of visibility for all ground stations of the system for a sufficiently long time. This requirement can be satisfied either with a very large number of uncontrolled satellites, or with a much smaller number of satellites, the position of which in orbit can be controlled.

Satellites which perform a whole number of rotations per stellar day and which have identical orbital parameters but are shifted in space and time so that the points beneath the satellites describe the same trajectory in sequence at equal time intervals are most convenient in this respect. They greatly facilitate switching and the operation of the communication system in general in comparison to uncontrolled satellites.

However, even for these periodic satellites, the continuity of the signal is disrupted at the moment of switching. Actually, since the two satellites are at different points in space at the moment of switching, the lengths of the paths between the two stations are different and a portion of the signal is either lost or repeated at the moment of switching.

A numerical estimate of this phenomenon can be produced on the basis of the information presented in the preceding section. Calculations show that the signal lost (or repeated) in systems with periodic satellites generally does not exceed 10-20 msec, while in systems with uncontrolled satellites this quantity may be on the order of 30 msec or more.

We should note in conclusion that in systems with periodic satellites, the use of which is most probable, the moments of switching can be easily predicted and follow each other at fully defined time intervals of several hours. If the requirements for signal continuity are rigid, variable delay devices can be used in the signal channel to equalize the travel time of the signal over the two paths before switching.

§ 4.6. Analysis of Possible Types of Communications Satellite Orbits

It is most expedient for communications satellites to use both circular and elliptical orbits with inclinations varying from 0 (equatorial orbits) to 90° (polar orbits), and orbital periods which are equal to or fractions of a stellar day. The use of orbits with inclinations of over 90° is less desirable due to the reduction in time of visibility as the satellite moves in the direction opposite to the daily rotation of the Earth. Furthermore, in this case the satellite will be launched in the direction opposite the daily rotation of the Earth, requiring the expenditure of additional energy (consumption of additional fuel) by the booster. /124

In selecting an orbital inclination, calculations are based primarily on the areas (latitudes) on the Earth which must be provided with communications. For example, equatorial orbits are suitable primarily for communications between points in the equatorial area, inclined orbits are suitable for points in the

low and middle latitudes, while polar orbits (generally circular) are suitable for global communications.

The possibility of producing an orbit with selected inclination by direct insertion of the satellite into the plane of the orbit is determined by the geographic latitude of the launch area. If the launch area is located at latitude ϕ, direct launch can be performed into any orbit inclined from the plane of the equator at angles of between ϕ and 90°. The greater the inclination, the higher the expenditure of energy required to insert the satellite (with launch from a given launch area), which is related to the greater variation between the direction of launch (launch azimuth) and the direction of daily rotation of the Earth. In order to insert a satellite into an orbit with an inclination of less than ϕ, including a stationary orbit, the satellite must be launched into an intermediate orbit with inclination ϕ, then the inclination of the orbit changed to the required value by subsequent maneuvering, requiring additional expenditures of energy for insertion.

The requirement of stability of the lines of apsides of the orbit, which determines the constancy of the area of servicing and communications conditions, has a significant influence on the selection of the inclination of elliptical orbits. The lines of apsides of an orbit will be immobile if an orbital inclination of near 63.4° is selected, eliminating the influence of the polar flattening of the Earth on rotation of the line of apsides. Adjustment of the position of the line of apsides after it drifts from its initial position involves great technical difficulties and great expenditures of power by the on-board orbital correction motors if the inclination of the orbit differs greatly from 63.4°.

In the Soviet Union, the Molniya-1 communications satellites are launched into orbits with an inclination of 65°, in which case the position of the line of apsides during the time of active existence of the satellites remains practically unchanged. In selecting the remaining orbital elements, such as perigee argument ω and longitude of ascending orbital mode Ω, counted from the direction to the point of the vernal equinox and related to the launch time of the satellite, calculations must be based on the fact that satellites must not enter the denser layers of the atmosphere during evolution, i.e., must continue orbital existence during the necessary time of functioning. /125

The use of an orbital period equal to or a multiple of the stellar day considering the evolution of parameter Ω during a rotation is dictated by the necessity of providing communications each day for the same regions of the Earth, which is possible only with a stable satellite track on the surface of the Earth. We must recall that as the orbital period increases, the altitude of the satellite increases, which expands the servicing zone,

increases the time of visibility of the satellite (communications session length). However, as altitude increases, the power expenditures (fuel consumption) of the booster used to insert it in orbit increase and, furthermore, higher power is required on the communications line, the ground and satellite electronic equipment becomes more complex.

As the flying altitude of the satellite and distance between ground points in communication increase, significant delays of signals, undesirable for duplex telephone conversations, appear, along with noticeable echo signals.

The optimal range of altitudes for communications satellites is 15,000-40,000 km.

Let us compare elliptical and circular orbits. Analysis of the dimensions of zones and times of visibility of satellites shows that with identical orbital periods, a satellite on an elliptical orbit with a high apogee can "illuminate" a given territory for a longer period of time than a satellite in a circular orbit (except for a stationary orbit).

As concerns the Doppler effect, the elliptical orbit, naturally, is less convenient than a circular orbit. However, sectors of an elliptical orbit near the apogee can be selected for communications, for which the Doppler frequency shifts are slight. As concerns the conditions and degree of technical complexity of launching of satellites, elliptical orbits with high apogees must be given preference, since with identical booster power, satellites carrying approximately twice the payload can be put in elliptical orbits as in circular orbits with the same period of rotation.

Satellites in stationary orbit (with periodic correction of the drift of the satellite caused by the equatorial flattening of the Earth) must be considered most convenient for communications purposes, for the following reasons:

1. One satellite can provide continuous, round-the-clock communications for almost one third of the Earth's surface.

2. Simplified, less expensive ground-station antennas and transceiving apparatus can be used.

3. Since the satellite is practically constantly in the field of vision of ground points, switching of ground station antennas from satellite to satellite is unnecessary and the number of ground station antennas and transceiver apparatus is halved in comparison with other types of orbits. /126

4. Round-the-clock servicing of a significant territory by a single satellite with simultaneous reduction in the volume of

ground station transceiver equipment allows less expensive systems to be created than systems using other types of orbits.

5. The practical absence of carrier frequency Doppler shift, the high stability and quality of communications due to the constant signal level at the inputs of receivers.

6. At an altitude of about 36,000 km, the satellite is significantly less influenced by the radiation belts than at lower orbits, which should increase the service life of the electronic equipment and solar batteries.

However, insertion of satellites into stationary orbits from launch areas not on the equator, the high orbital altitude and the necessity for elimination of drift of a stationary satellite represent very complex technical problems.

REFERENCES

4.1. Shtern, T., Vvedeniye v nebesnuyu mekhaniku [Introduction to Celestial Mechanics], Mir Press, 1964.

4.2. El'yasberg, P. Ye., Vvedeniye v teoriyu poleta iskusstvennykh sputnikov Zemli. [Introduction to the Theory of the Flight of Artificial Earth Satellites], Nauka Press, 1965.

4.3. Perkins, "Flight Mechanics of a Stationary Satellite," Raketnyye tekhnika i kosmonavtika [Rocket Equipment and Astronautics -- Collection of Works], No. 4, 1963.

CHAPTER 5. FREQUENCY RANGES USED FOR COMMUNICATIONS BY SATELLITE AND PECULIARITIES OF THEIR USE

§ 5.1. Factors Influencing the Selection of Frequency Ranges /127

One of the basic difficulties involved in the development of space communications continues to be the limitation of weight and electric power consumption of equipment installed on satellites.

Considering these limitations, satellite communications systems are currently planned so that the satellite relay station is a combination of a simple receiver device with a reliable, low-power transmitter, while the ground stations use large antennas, powerful transmitters and highly sensitive receivers. Proper selection of the operating frequency range in this case goes a long way toward determining the complexity and cost of the ground stations and, consequently, of the entire communications system.

In selecting frequency ranges, one must first of all keep in mind that the influence of the surrounding medium at various frequencies may result in significant attenuation of signals transmitted in the satellite communications system and in the development of excessive noise, produced by natural and artifical sources.

The signals transmitted from a satellite to the Earth or sent from the Earth to a satellite pass through the entire thickness of the atmosphere and it is this interaction with the atmosphere which causes their attenuation to some extent.

The terrestrial atmosphere has frequency selectivity, manifested in that electromagnetic oscillations at certain frequencies pass through it almost without losses, whereas it is "nontransparent" for oscillations at other frequencies.

The atmosphere is "transparent" basically within the following two frequency ranges:

-- the frequency range between the ionospheric critical frequencies and frequencies of absorption of atmosphere precipitation and gases (approximately from 10 MHz to 20 GHz);

-- the range of visible and infrared rays.

The first of these two ranges has been most fully used and is /128 therefore of greatest interest. All contemporary experimental and operational communications satellites, as well as the tracking, telemetry and remote control systems servicing them, operate within this range.

We cannot state the precise boundaries of this range, since they depend on many factors. For example, the upper boundary of the range may fall below 10 GHz during strong rain, the lower boundary may rise to 70 MHz, depending on the state of the ionosphere (degree of solar activity), the geographic position of the ground station and the geometry of the signal propagation path. On the other hand, the boundaries of the range may extend down to 2 MHz for the polar regions at night and up to 50 GHz for high mountain regions with dry air.

When frequencies near the boundaries of the range are used, the influence of interfering radiation from natural sources of terrestrial and extraterrestrial origin and industrial noise must be considered.

Near the lower boundary, the noise of natural extraterrestrial sources ("cosmic noise") becomes significant. This noise originates from three types of sources: the galactic background radiation, point source radiation and sun noise.

Point sources and the sun, due to their small angular dimensions, rarely fall within the main lobe of the radiation pattern of an antenna and need not be considered in selecting a frequency range. As concerns the galactic background radiation, its level decreases with increasing frequency and at frequencies over 1 GHz it becomes insignificant in comparison to the natural thermal noise of the receiver.

Near the upper boundary of the range, the signal/noise ratio at the input of the receiver may deteriorate due to the thermal radiation of the atmosphere of the Earth, which increases sharply at frequencies over 10 GHz.

Among other atmospheric effects influencing the transmission of signals, we must note the Faraday effect and the refraction of a radio beam in the atmosphere. Refraction causes bending of a radio beam as it passes through the layers of the atmosphere. The bending which occurs depends little on frequency and, therefore, cannot be greatly reduced by frequency selection.

The Faraday effect consists in rotation of the plane of polarization of a linearly polarized planar wave as it passes through the ionosphere in the presence of the magnetic field of the Earth. The influence of the Faraday effect can be eliminated by using circularly polarized waves, which is also required by the unavoidable randomness of the position of the satellite antenna relative to the ground station antennas.

Thus, considering all of the propagation factors noted above[1], it has been concluded that from the standpoint of propagation

[1]Detailed quantitative analysis of all of these factors is presented in Chapter 7.

conditions, frequencies in the 1-10 GHz band are best for communications lines between ground stations and satellites. /129

However, by the time the first communications satellites were launched, the 1-10 GHz band was already intensively used by various services and the insertion of a new service into this band -- space communications, with its powerful transmitters, large antennas and sensitive receivers -- would have resulted in mutual interference.

The probability of mutual interference, magnitude of interference and degree of interfering action when two or more services operate in the same frequency band depend on the nature of the services and the technical parameters of the electronic devices used.

The problem of assignment of specific frequency bands for space communications can be reduced, therefore, first of all to determination of services operating in the 1-10 GHz band which would be most compatible from the standpoint of mutual interference and, secondly, to development of measures toward decreasing the degree and probability of mutual interference which would not limit further development of these services operating in the same frequency band.

The rapid development of space communications and the prospects which have been opened for its use have required that frequencies be assigned for space communications as rapidly as possible, in order to allow effective planning of space communications systems.

§ 5.2. Allocation of Frequency Bands for Space Communications

Problems of the allocation of frequency bands are being studied by the International Telecommunications Union (ITU) which, on the basis of studies performed in various countries and presented to the International Consultative Committee for Radio (ICCR) develops recommendations and rules controlling the operation of telecommunications services throughout the world at its conferences.

One administrative conference on telecommunications (Geneva, 1959), in its Recommendation No. 36, recommended that the administrative council of the ITU call an Extraordinary Administrative Conference on Telecommunications (EACT) in 1963 to allocate frequency bands for space communications services.

The most important result of the work of the EACT is the revision of Article 5 of the telecommunications regulations (Geneva, 1959), reflecting the distribution of frequencies by services and zones.

The world is divided into three frequency band distribution zones: /130

Region 1 -- Europe, Africa, USSR, Mongolia;
Region 2 -- N. and S. America, Greenland;
Region 3 -- Asia (excluding the Asiatic portion of the USSR and Mongolia), Australia (for more detail, see map in Telecommunications Regulations, Geneva, 1959).

According to the decisions of the EACT, satellite communications services are allocated the following frequency bands (frequencies indicated in MHz):

Region 1	Region 2	Region 3
3400-4200 (Space-Earth)	3400-4200 (Space-Earth)	3400-4200 (Space-Earth)
4400-4700 (Earth-Space)	4400-4700 (Earth-Space)	4400-4700 (Earth-Space)
5725-5850 (Earth-Space)	---	---
5850-5925 (Earth-Space)	---	5850-5925 (Earth-Space)
5925-6425 (Earth-Space)	5925-6425 (Earth-Space)	5925-6425 (Earth-Space)
7250-7750 (Space-Earth)	7250-7750 (Space-Earth)	7250-7750 (Space-Earth)
7900-8400 (Earth-Space)	7900-8400 (Earth-Space)	7900-8400 (Earth-Space)

The frequency bands allocated for transmission from space to Earth can also be used for transmission of telemetry and tracking signals, while the frequency bands allocated for transmission from Earth to space can be used for transmission of remote control signals from ground stations.

The frequency bands indicated in the table are designed for use in communications systems using active relay satellites. As concerns systems with passive satellites, the EACT agreed that these systems can be placed as an exception in the 7250-7750 MHz band on agreement between interested administrations and administrations whose services operating in accordance with the telecommunications regulations might be interfered with, and also if all limitations developed by the EACT for communications systems using active relay satellites are followed.

§ 5.3. Limitations Placed on Communications Services Using Satellites and Other Services Working in the 1-10 GHZ Band

/131

The bands of frequencies allocated by the EACT for space communications must be used by this service jointly and with equal rights with fixed or mobile communications stations.

In order to avoid excessive mutual interference and to assure normal development of all services working together in the frequency band, the EACT has also developed certain limitations on the technical parameters, locations and order of operation of stations of these services. In general, these limitations can be reduced to the following main requirements, designed to avoid interference for all services:

-- placement of transmitting stations and, where the nature of the service allows, placement of receiving stations should be selected with particular care;

-- radiation in unnecessary directions and reception from these directions must be minimized if the nature of the service allows it, by maximum utilization of the directional properties of antennas;

-- selection and utilization of transmitters and receivers must be performed according to the statements of the telecommunications regulations (for example, norms for maximum power, lateral radiation, frequency deviation, band occupied, etc.);

-- space stations must be able to stop radiating during those times when they are not actually in use for communications (for example, in the intervals between communications sessions or after completion of service life).

Based on specific analysis of possible cases of mutual interference, the EACT also developed the following technical limitations, placed on the space communications service and on services working together with it in the same frequency bands[1]:

1. Maximum equivalent isotropically radiated power[2] of transmitter and related antenna for a fixed or mobile station should not exceed +55 dbw.

2. The power fed to the antenna by the transmitter of a fixed or mobile station should not exceed +13 dbw, i.e. 20 w.

---

[1]The limitations presented below are preliminary and subject to further study.

[2]The maximum equivalent isotropically radiated power of a transmitter is defined as the product of the power supplied by the transmitter to the antenna times the gain of the antenna relative to an isotropic radiator.

The limitations outlined in 1 and 2 above apply to the following frequency bands, selected for reception of satellite signals, used jointly with fixed or mobile services: /132

5800-5850 MHz (in Albania, Bulgaria, Hungary, Poland, Rumania, Czechoslovakia and the USSR);
5850-5925 MHz (regions 1 and 3);
5925-6425 MHz;
7900-8100 MHz.

In order to protect communcations satellites in stationary orbit from interference due to line-of-sight radio relay systems using the frequency band shared by satellite receivers, the 11th Plenary Assembly of the ICCR (Oslo, 1966) developed a recommendation, in which it was suggested that the new paths of these radio relay lines be planned so that the axis of the main lobe of the antenna be directed at least 2° away from the orbits of stationary satellites.

If in any specific case this limitation cannot be observed, all efforts should be expended to limit the equivalent isotropically radiated power of the radio relay line transmitter to the following maximum values:

+47 dbw for antennas directed less than 0.5° away from a stationary orbit;

from +47 to +55 dbw (with linear change of 8 db per degree) for antennas directed between 0.5 and 1.5° away from a stationary orbit.

These requirements can be rather easily fulfilled in planning a new radio relay line, causing no significant increases in total costs.

In the case of creation of a radio relay system based on existing lines, the maximum equivalent isotropically radiated transmitter power, when possible, should not exceed these values, but only for points in the stationary orbit announced internationally.

3. The mean equivalent power isotropically radiated by a ground station in satellite communications services in any direction in the horizontal plane[1] should not exceed +55 dbw in any band 4 KHz wide.

4. In any direction where the distance from the ground station to the boundary of the territory of another administration is over 400 km, the limit of +55 dbw in any 4 KHz band for this direction can be increased by 2 db for each 100 km of distance after the first 400 km.

[1]The true value of radiated power in each specific case should be decreased by a factor considering the shielding value of surrounding terrain relief.

This limit can be exceeded only by agreement between the interested administrations, but in no case can the radiated power as defined in paragraph 3 exceed +65 dbw in any 4 KHz band. /133

The limitations outlined in paragraphs 3 and 4 are used in the following frequency bands, allocated for transmission from ground stations to communications satellites and jointly used with equal rights by fixed or mobile communications services:

4400-4700 MHz;
5800-5850 MHz (in Albania, Bulgaria, Hungary, Poland, Rumania, Czechoslovakia and the USSR);
5350-5925 MHz (regions 1 and 3);
5925-6425 MHz;
7900-8400 MHz.

5. The antennas of ground transmitting stations in communications satellite systems should not be used at elevation angles of less than 3° measured from the horizontal plane to the central axis of the main lobe of the antenna, unless this has been coordinated between the interested administrations.

This limitation is applicable in the following frequency bands, allocated for transmission from ground stations to communications satellites, when these frequencies are jointly utilized with equal rights by fixed or mobile services;

4400-4700 MHz;
5800-5850 MHz (Albania, Bulgaria, Hungary, Poland, Rumania, Czechoslovakia and the USSR);
5850-5925 MHz (regions 1 and 3);
5925-6425 MHz;
7250-7750 MHz;
7900-8400 MHz.

6. In order to decrease interference to surface services by satellite transmitters, the EACT has developed the following limitations:

-- the power flux density at the surface of the Earth, created by the radiation of the transmitter on a communications satellite or by reflection from a passive communications satellite, utilizing angular (frequency or phase) modulation with great frequency deviation, should not exceed -130 dbw/m^2 for all angles of arrival. Furthermore, these signals, if necessary, should be continually modulated by a signal of the proper shape so that the power flux density does not exceed -149 dbw/m^2 in any 4 KHz band for all angles of arrival;

-- when other methods of modulation are used, the power flux density created at the surface of the Earth should not exceed -152 dbw/m^2 in any 4 KHz band for all angles of arrival.

The limitations outlined in section 6 are used in the following frequency bands, allocated for transmission from communications satellites and jointly used with equal rights by fixed or mobile /134

services:
3400-4200 MHz;
7250-7750 MHz.

These limitations were developed on the basis of studies performed prior to 1963 and, since these studies were based on insufficiently complete information, contradictions were quickly discovered in the very nature of the limitations and new information was developed, allowing them to be refined.

For example, in particular, studies of possible paths of development of interference for line-of-sight radio relay lines by Earth satellite transmitters and approximate calculations of the interference which could arise at the output of such radio relay lines allowed a recommendation to be developed at the 11th Plenary Assembly of the ICCR (Oslo, 1966), in which it is suggested that in the frequency bands used jointly by satellite communications systems and radio relay lines, the radiation flux density created at the surface of the Earth by the radiation of transmitters on active satellites under all conditions and with all modulation methods not exceed $(-152 + (\beta/15))$ dbw/m^2 in any 4 KHz band[1]. Here β is the angle of arrival of the satellite signal, read in degrees from the horizontal plane.

Thus, the radiation flux density created by a satellite transmitter in the horizontal direction, where there may be radio relay lines with antennas pointing nearly at the interfering satellite is more sharply limited, since in all other cases (with greater angles of elevation of the satellite relative to radio relay receivers), the signal received from the satellite will be significantly attenuated due to the selective properties of the reception pattern of the radio relay station antenna.

7. Space stations must be equipped with devices for rapid interruption of their radiation when necessary (for example, upon receipt of instructions from the Earth, by timing devices on board the satellite, due to the operating life of the power supplies, etc.).

§ 5.4. Coordination of Assignment of Frequencies to Space and Ground Service Stations

The limitations listed in the preceding section cannot fully solve the problem of mutual interference.

For example, even with these limitations, it may occur that a ground station of a satellite communications service and a

[1]The values indicated in section 6 relate to the radiation flux density calculated for the propagation conditions in free space.

surface communications station are located so close together that /135
the radiation of one station significantly exceeds the signal level received by the other station. Naturally, normal operation of the stations under these circumstances is impossible.

To avoid such cases, the EACT has developed a special order of assignment of frequencies to staticns involved in space and surface communications services operating in the same frequency band.

According to this order, under certain conditions administrations should report every allocation of a frequency to a surface or space station to the International Committee for Registration of Frequencies (ICRF). However, before the request is sent to the ICRF, the order for assignment of frequencies developed by EACT requires that the administrations go through a so-called "coordination procedure," the purpose of which is to determine whether a planned station will serve as a source of interference for existing nearby stations or whether the planned station will be interfered with by these existing stations.

The coordination procedure begins with calculation of a "coordination distance" for the planned location of the station.

The coordination distance in a given direction from a receiver station is defined as that distance, beyond which a transmitting station from another service could be located without causing excessive interference with the planned receiving station with its known technical parameters. The coordination distance in this direction from a transmitting station is respectively defined as the distance, beyond which the receiving station of another service could be located without experiencing excessive interference from the planned transmitting station with its known technical parameters. The coordination distance is defined on the basis of the known parameters of the planned station, the supposed parameters of stations of other services and data on propagation for the local area. The parameters of the station of the other service should be selected with some reserve factor in order to decrease the probability of error caused by the unavoidable variations in these parameters encountered in practice.

In order to facilitate calculations and avoid various misunderstandings, EACT has confirmed a temporary method for calculation of the coordination distance, which will be described below. The calculation of the coordination distance in all azimuthal directions from a given station allows a zone to be constructed, within which stations of another service might create interference for the planned station or might be interfered with by it. If the territory of one or more member countries of the ITU falls within this zone, the administrations of these countries are sent copies of a map in a suitable scale indicating the location of the planned station and the coordination zone, as well as information on the /136
primary parameters of the station, based on which the zone was

constructed, indicating the approximate date when operations are planned to begin.

Based on this information, the administrations evaluate the possibility of mutual interference with the planned station and, if necessary, seek means to solve the problem with the administration which originated the coordination.

If no other member countries of the ITU fall within the coordination zone, the coordination procedure is not required and the location and parameters of the stations are reported to the ICRF with an indication that coordination is not required.

Coordination is also not required if the planned ground station is designed to operate in the 4400-4700 MHz or 8100-8400 MHz band or if it will be installed on board a ship or aircraft. In the latter case, however, if a station installed on board a ship or aircraft falls within the limits of the coordination distance in relationship to the boundaries of another country, its operation must be coordinated with the administration of this country, regardless of the frequency within the 1-10 GHz band used by the station.

§ 5.5. Calculation of the Coordination Distance

The coordination distance must be most frequently determined in the case of possible interference between the ground station transmitter and receivers of line-of-sight radio relay lines or a line-of-sight radio relay line transmitter and the ground station receiver of a satellite communications system.

In the calculations, the method of which is outlined in this section, a high gain of the receiving station antenna is always used in order to produce a clearly greater coordination distance than actually necessary. For this same reason, it is always assumed that sensitive receivers with low noise level are used.

Calculation of the coordination distances is reduced to determination of the necessary minimum permissible primary attenuation which the noise must experience as it propagates, using the following formula:

$$L_0 = (P_i + G_i) - F - (P_r - G_r), \qquad (5.1)$$

where P_i is the power of the interfering transmitter at the input of the antenna-feeder device, dbw;

G_i is the gain of the transmitting antenna of the interfering station in the direction toward the receiving station subject to the interference, calculated for an isotropic antenna considering

feeder losses, db;

F is the shielding factor of the terrain around the ground station, db; /137

P_r is the maximum permissible noise level at the input of the receiver of the receiving station, dbw;

G_r is the gain of the antenna of the receiving station in the direction toward the interfering transmitter, calculated for an isotropic antenna and decreased by the losses in the feeder, as well as the losses caused by differences in polarization if any, db.

In analyzing the interference created by telephone transmission systems, particularly systems using frequency modulation, it is convenient to perform calculations on the basis of the power density in the 4 KHz band. Therefore, in the case of interference from a ground station transmitter affecting surface radio relay systems, the value of P_i is taken as the maximum power density permissible in any 4 KHz band fed from the transmitter of the ground station to the input of the feeder, and similarly P_r is the maximum permissible power density at the input of the receiver for any 4 KHz band.

In studying noise from a surface relay line transmitter interfering with the receiver of a ground station, the powers P_i and P_r in equation (5.1) can be more conveniently calculated as the total powers, not the power density.

This differentiated approach to the two systems is explained by the fact that in satellite communications systems, it is assumed that FM will be used with such great frequency deviation that only a portion of the signal spectrum radiated by the satellite system will be within the transmission band of the radio relay station receiver, so that the maximum spectral power density is more important than the total power of the interfering signal. When satellite systems are interfered with by radio relay stations, the interfering signal may fall fully within the frequency band of the receiver and therefore it is the total power of the interfering signal which is important.

Furthermore, it was assumed that in a satellite communications system in the case of slight loading, the carrier is modulated by a signal of a special form which scatters the signal energy over a broad frequency band.

The minimum permissible primary attenuation during propagation in any direction from the transmitting station is determined on the basis of the following considerations:

First Case. Interference from ground station transmitter of satellite communications service to surface line-of-sight radio relay system.

The noise arising in a telephone channel at the output of the radio relay line as a result of interference by the transmitting stations of communications satellite services is limited by Recommendation 357 of the ICCR to 100,000 pw (-40 dbm 0)[1] for not more than 0.01% of the time. /138

Considering that up to 4 non-simultaneous noise sources may be in operation over the entire length of a radio relay line, this norm for interference from a signal source (interfering ground station) should not be exceeded more than 0.0025% of the time.

Assuming that the interfering station is lightly loaded and that some method of scattering of the energy of the carrier is used so that the spectrum of the interfering signal is distributed evenly over a band of at least 300 KHz, we can show[2] that interference of this type is attenuated during transmission through the receiver channel of the radio relay station by not more than 1 db[3].

Consequently, the maximum ratio of interfering signal to useful signal at the input of the receiver should not exceed -39 db more than 0.0025% of the time.

Since the minimum level of useful signal to input of the receiver of a typical radio relay station is -74 dbw[4], the permissible level of interfering signal at the input under all of the assumptions made above should not exceed -113 dbw more than 0.0025% of the time.

The permissible interfering signal level at the input of the receiver in any 4 KHz band therefore should not exceed -132 dbw.

---

[1]This recommendation was modified at the 11th Plenary Assembly of the ICCR, Oslo, 1966 (see Chapter 10), the noise power noted here relates to unweighted noise in a 4 KHz band.

[2]Using methods presented or mentioned in Chapter 10.

[3]This quantity was taken from the example of a radio relay system using line-of-sight transmission and carrying 960 channels.

[4]This quantity is also taken from the example of a line-of-sight 960 channel radio relay system, the maximum permissible interfering signal level -- 113 dbw -- being almost independent of the number of channels.

Further, assuming that the gain of the radio relay station antenna relative to an isotripic radiator, considering losses in the feeder, is 42 db[1], the gain of the ground station antenna in the horizontal plane relative to an isotropic radiator considering losses in the feeder and possible losses due to differences in polarization is G_i (db) for 2.5% of the time and that the power fed by the ground station transmitter to the input of the feeder in a 4 KHz band is P_i (dbw), we find that the minimum permissible primary attenuation in propagation for at least 0.1% of the time should be no less than

$$L_0 = P_i + G_i - F + 174, \text{ db}, \qquad (5.2)$$

where F is the shielding factor of the terrain for the ground station (if present).

For simplicity, we can assume G_i to be the maximum value /139 obtained in the horizontal plane in the corresponding azimuthal direction, rather than the value exceeded 2.5% of the time. However, if terrain shielding occurs, the value used should be the maximum value produced with an elevation corresponding to the shielding obstacle.

Second Case. Interference from transmitter of surface line-of-sight radio relay station to ground station of communications satellite system.

The noise arising in a telephone channel at the output of the communications system resulting from interference by surface radio relay stations is limited by Recommendation 556 of the ICCR to 160,000 pw (-38 dbm 0)[2] for not over 0.02% of the time.

It was assumed that one ground station might be interfered with by four radio relay stations, not simultaneously, so that this norm for interference from one radio relay station should not be exceeded more than 0.005% of the time.

Based on the most probable parameters of the interfering signal and assuming that the satellite communications system uses some method of spreading the carrier, it can be shown[3] that noise of this type is attenuated upon transmission through the receiving system of the ground station by not over 10 db[4].

---

[1] If we know that the gain exceeds 42 db, the true figure should be used.

[2] This Recommendation was revised in the 11th Plenary Assembly of the ICCR, Oslo, 1966 (see Chapter 10). The noise power indicated here relates to unweighted noise in a 4 KHz band.

Consequently, the maximum ratio of interfering signal to useful signal at the input of the receiver should not exceed -28 db more than 0.005% of the time.

Since the minimum level of useful signal at the input of the receiver of a typical ground station of a satellite communications system is -117 dbw[4], the permissible level of the interfering signal of the input of a receiver, under all the assumptions made above, should not exceed -145 dbw more than 0.005% of the time.

Further, assuming that the gain of the antenna system of the ground station in the horizontal plane relative to an isotropic radiator is G_r (db) for 5% of the time, that the gain of the radio relay station antenna relative to an isotropic radiator, considering losses in the feeder, is G_i (db) and that the power fed by the transmitter of the radio relay line to the input of the feeder is P_i (dbw), we find that the minimum permissible primary attenuation upon propagation for at least 0.1% of the time should be no less than /140

$$L_0 = P_i + G_i - F + G_r + 145, \text{ db}, \qquad (5.3)$$

where F is the terrain shielding factor for the ground station (if there is shielding).

For simplicity, we can take as G_r the maximum value produced in the horizontal plane in the corresponding azimuthal direction, rather than the value exceeded 5% of the time. However, if terrain shielding is present, the quantity used should be the maximum at the angle of elevation corresponding to the shielding obstacle.

In those cases when ground stations are located below the level of the surrounding terrain, some additional attenuation caused by the shielding influence of the surrounding terrain must be considered. For example, if an obstacle in the given azimuthal direction has an elevation of β relative to the ground station, in calculating the coordination distance in this direction the maximum gain of the antenna of this ground station with an elevation of the antenna of β must be used, not the maximum gain with the antenna laid horizontal.

---

[3]Using method presented or mentioned in Chapter 10.

[4]This value was taken from the example of a satellite communications system with 1200 channels, the maximum permissible interfering signal level of -145 dbw being almost independent of the number of channels.

As was noted above, where terrain shielding occurs, the necessary primary attenuation in propagation may be reduced by the shielding factor F, expressed in decibels.

When an obstacle limiting the elevation is located at over 5 km from the ground station, the terrain shielding factor shown in Table 5.1 should be used.

TABLE 5.1

Minimum Elevation of Obstacle Visible from Ground Station (β)	Below 1°	Between 1° and 2°	Between 2° and 3°	Between 3° and 4°	Between 4° and 5°	Over 5°
Permissible Terrain Shielding Factor (F), db	0	5	8	11	13	15

If obstacles are located closer, the terrain shielding factors can be produced by multiplying those presented in Table 5 by $\ell/5$, where ℓ is the distance from the ground station to the obstacle in kilometers.

The values of terrain shielding factor presented here should be used carefully if ground stations are located in areas within the coordination distance above the horizontal plane passing through the ground station. /141

The propagation data presented below relate to a frequency of 4 GHz. Therefore, before using these data to determine the coordination distance at any frequency in the 1-10 GHz band, the minimum permissible primary attenuation during propagation (L_0) must be converted to the equivalent attenuation at 4 GHz (L'_0). The equivalent attenuation in decibels at 4 GHz is determined by the formula

$$L'_0 = L_0 + 13 - 21.6 \log_{10} f, \qquad (5.4)$$

where f is the assigned frequency, GHz.

This relationship is also illustrated graphically by Figure 5.1, which shows the dependence of the correcting factor which should be added to the necessary attenuation at frequency f to produce the equivalent attenuation at frequency 4 GHz.

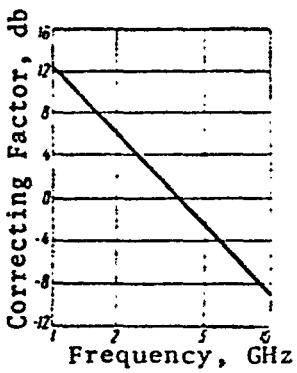


Figure 5.1. Correcting Factor Which Should be Added to Necessary Attenuation L_0 at Frequency f to Produce Equivalent Attenuation L'_0 at 4 GHz

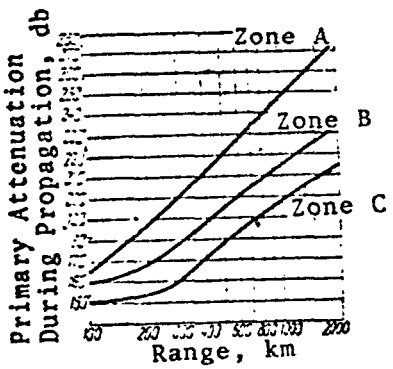


Figure 5.2. Simplified Tropospheric Scatter Curves for Calculation of Coordination Distance. Primary Attenuation during Propagation not Exceeded for Over 0.1% of Time at 4 GHz.

The value of L'_0 can be determined as a function of distance using Figure 5.2. The propagation curves presented in Figure 5.2 relate to zones A, B and C and correspond to the following basic climatic telecommunications area on the Earth: /142

Zone A: Continental;

Zone B: At sea, latitudes over 23.5° n. lat. and 23.5° s. lat.;

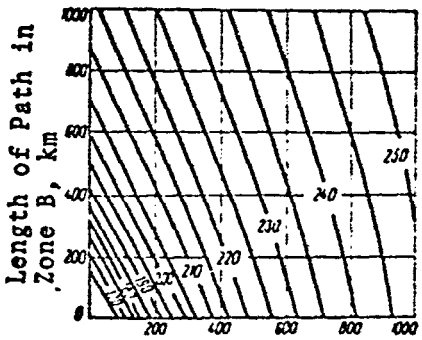
Zone C: At sea, latitudes between 23.5° n. lat. and 23.5° s. lat.

In any direction from the ground station, the necessary coordination distance is determined as follows.

a) if the primary equivalent attenuation during propagation is such that the coordination distance in a given direction falls fully within the limits of one of the zones, this coordination distance can be determined directly from Figure 5.2, using the corresponding curves;

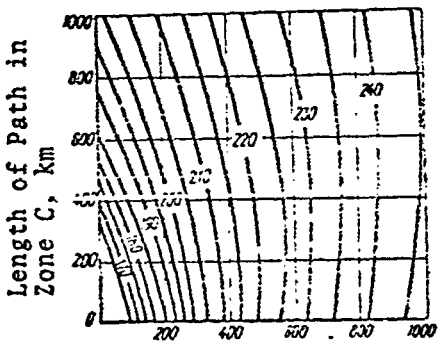
b) if the coordination distance falls partially within one zone and partially within another, curves for mixed paths (Figures 5.3, 5.4 and 5.5) should be used. These curves show the equivalent attenuation as a function of length of the path in each of the two zones individually, namely: Figure 5.3 shows zones A and B; Figure 5.4 shows zones A and C; Figure 5.5 shows zones B and C.

Thus, if we know the length of a path in one zone and the necessary attenuation, we can determine the length of the path in the other zone. The path length in the first zone is the known distance from the ground station to the boundary of the zone in the corresponding direction; consequently, the path length in the other zone is easily determined.



Length of Path in Zone A, km

Figure 5.3. Graph for Calculation of Coordination Distances -- Mixed Paths in Zones A and B. Primary Attenuation During Propagation not Exceeded for 0.1% of the Time at 4 GHz.



Length of Path in Zone A, km

Figure 5.4. Graph for Calculation of Coordination Distances -- Mixed Paths in Zones A and C. Primary Attenuation During Propagation not Exceeded for 0.1% of the Time at 4 GHz.

The total path length, or coordination distance, is naturally the sum of the two path lengths. /143

An example of calculation of coordination distance for a mixed path is presented below;

c) in certain geographic areas, where attenuation is less than the values calculated using the zonal propagation curves, coordination distances must be determined on the basis of known propagation data for these areas.

The method used in the case of a mixed path is illustrated by the following example, in which it is assumed that in order to avoid interference from a ground station to surface communications services, a primary equivalent attenuation of 190 db is necessary in the direction in question.

As we can see from Figure 5.6, the ground station is located at 50 km from the shore, followed by 150 km of water. The coordination distance from the ground station in this direction must then be determined, using the propagation graph for mixed paths, Figure 5.7. Calculations are performed as follows:

1. First, the distance from the ground station to the shoreline (50 km) is laid out on the A axis of a graph, as shown by point A_1. /144

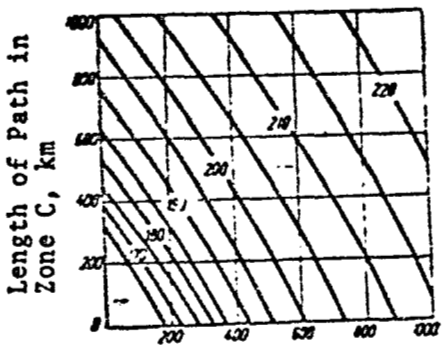


Figure 5.5. Graph for Calculation of Coordination Distances -- Mixed Paths in Zones B and C. Primary Attenuation During Propagation not Exceeded for 0.1% of the Time at 4 GHz.

2. Then, the length of the path over the sea (150 km) is laid out on the B axis of the graph, as shown by point B_1.

3. The desired distance over the land is measured parallel to the A axis from point B_1 to the point of intersection with the 190 db curve, marked as X. We find that this distance is 90 km.

4. The coordination distance is therefore the sum of the coordinates of point X and is equal to 50 km + 150 km + 90 km = = 290 km.

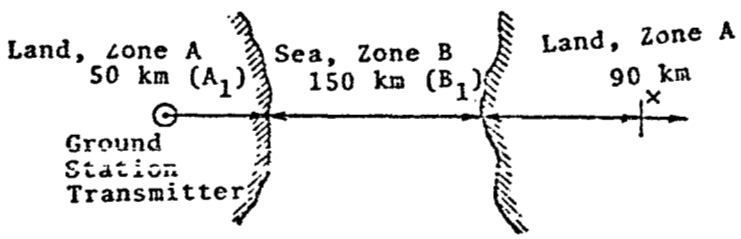


Figure 5.6. Mixed Path for Which Coordination Distance is Calculated.

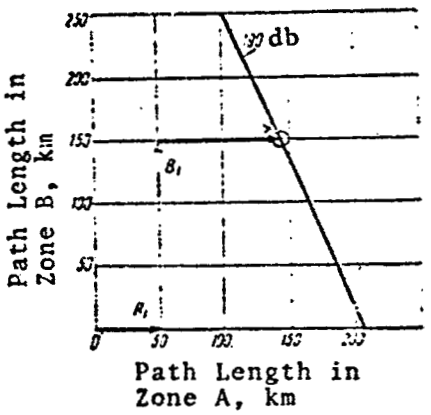


Figure 5.7. Calculation of Coordination Distances for Mixed Paths

CHAPTER 6. PERFORMANCE FACTORS OF SATELLITE COMMUNICATIONS CHANNELS

/145

§ 6.1. General Statements. Hypothetical Circuit

The requirements placed on communications channels generally should be independent of the actual physical means used to organize a channel; actually, the quality of the television image provided a receiver should be identically high, regardless of whether the image is transmitted over a cable line or by satellite. Single norms for permissible distortions of various types of transmitted messages actually exist and have been established both by international agreement and by internal Soviet Union standards, the international and union standards mostly agreeing. Nevertheless, specific peculiarities of the communications lines used by satellites and certain difficulties in the creation of these lines have forced the International Consultative Committee for Radio (CCIR) to adopt in some cases special recommendations for the performance factors of satellite communications lines. We will also indicate these peculiarities and attempt to estimate the magnitude of deviations from the standard norms, and whether they need by considered in planning communications systems. Before making the requirements for communications lines specific, it is desirable to define more precisely the type of line in question or, using the language of the International Recommendations, to establish the hypothetical standard circuit to which the norms established directly relate. In contrast to surface communications lines, it is not necessary to define the line length for satellites, since the parameters of a satellite communications line are almost independent of the distance between ground stations, measured over the surface of the Earth. The number of so-called "hops," i.e., Earth-satellite-Earth transmission paths, is very significant, however. Series connection of two or even three such hops is necessary in the creation of a global communications system. However, since the distance between Earth stations may be tremendous with a single step -- up to 18,000 km, most communications will be organized within a single step. This is true for intercontinental lines and virtually obligatory for intracontinental lines, including communications within the limits of the USSR. Recommendation 352 of the ICCR establishes that the hypothetical standard circuit for a communications system using an active relay satellite consists of one Earth-satellite-Earth line (see Figure 6.1), containing one modulator-demodulator pair, necessary for conversion of the message to a radio signal and back.

/146

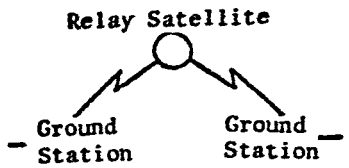


Figure 6.1. Basic Hypothetical Standard Circuit

§ 6.2. Performance Factors of Satellite Communications Channels for Transmission of Television

The most obvious feature distinguishing a television channel from channels for most other messages is its great bandwidth. The lower boundary of the band of frequencies is determined by the field frequency and is 50 Hz for the Soviet standard. The upper boundary, according to Recommendation 421 of the ICCR, may vary depending on the standard of the country involved, from 3 to 10 MHz. The Soviet standard calls for an upper boundary of 6 MHz. This band is sufficient for production of a sharp image corresponding to the number of lines used in the USSR, 625, and also for transmission of color television. This value should therefore be utilized, in spite of the fact that Recommendation 354 of the ICCR recommends an upper video frequency limit of 5 MHz for communications systems using active satellites transmitting 625-line television pictures. The ICCR recommendations base this limitation on the fact that the expenditures for construction and operation of satellite communications systems depend to a significant extent on the necessary video spectrum bandwidth, which "should not exceed values actually needed for transmission with acceptable quality." Actually, limiting the bandwidth to 5 MHz has little influence on image quality. Nevertheless, one should consider that expansion of the bandwidth to 6 MHz practically does not reduce the signal to weighted noise ratio, due to the slight role of the high frequency noise components, and therefore requires no increase in the energy potential of the radio line. Expansion of the video signal itself represents no significant difficulties. The need to transmit color television (with the 4.43 MHz color subcarrier) is a decisive argument in favor of planning future systems for a 50 Hz-6 MHz band. The desirability of expanding the band to 5.5-6 MHz /147 has been noted in several ICCR documents (1V/1006-E, Oslo, 1966).

The signal to noise ratio is the most important factor of a satellite communications line, having the greatest influence on the complexity and cost of the line.

Recommendation 421 of the ICCR, common for black and white long-range television transmission lines, establishes a norm of noises at not less than -57 db for the 6 MHz band, -52 db for the 5 MHz band during the time exceeding 1% of the time of any month; during time exceeding 0.1% (but not over 1%), noise can rise by not over 8 db. For satellite lines, Recommendation 354 of the ICCR allows a signal to noise ratio of -55 db for 99% of the time. Apparently, this value should be considered the limit toward which

we strive in planning satellite communications systems. The problem of the noise permissible for less than 1% of the time for satellite lines has not yet been resolved. However, we should recall that greater noise than - 55 db can be permitted for short periods of time, i.e., in designing a line according to this norm, one should not consider the simultaneous combination of the most unfavorable circumstances (strong rain, maximum error in orientation of all antennas, etc.).

The satellite line can generally be equated to a surface line 5,000-7,500 km long. With this length, domestic norms permit an increase in noise level to -52 - -54 db.

The method of measurement of the signal-to-noise ratio in order to compare it to the norm requires certain explanations. The signal is evaluated according to its amplitude (peak to peak), but without considering synchronizing pulses, i.e., 70% of its total amplitude. The video signal is evaluated according to amplitude due to the usual method of its measurement using a calibrated oscilloscope. In performing energy calculations for radio lines, the effective value of sinusoidal measurement signal is usually used, 9 db less than the peak-to-peak amplitude.

Noise is normalized according to its effective value, so that certain errors may arise when measurements are performed by a voltmeter calibrated for a sinusoidal signal. Weighted noise is measured through a weighting (visometric) filter, considering the reduced sensitivity of the vision to high frequency noise components. The attenuation introduced by this circuit is determined by the following formula (ICCR Recommendation 421)

$$\alpha = 10 \log [1 + (2\pi\tau F)^2], \qquad (1)$$

where F is the frequency,
τ = 0.33 μsec (with a bandwidth of 6 MHz).

With frequency modulation, normally used in satellite lines, the noise spectrum at the output of the receiver is triangular; therefore, the weighted filter greatly reduces noise -- by 17.8 db /148
in the 6 MHz band. Thus, the ratio of the effective value (or simply power) of the video signal to effective value (power) of noise, measured directly at the output of the frequency detector and usually produced in power calculations, should be only

$$55 - 9 - 17.8 = 28.2 \text{ db}.$$

If we have in mind modulation of the HF carrier by a sinusoidal signal with an amplitude equal to the full amplitude of

the video signal including synchronization pulses, we must provide

$$28.2 + 3 = 31.2 \text{ db.}$$

It should be recalled that the same weighting circuit decreases noise in the narrower video band to a lesser extent; for example, in a 5 MHz band, the gain from the use of the weighting filter is 16.3 db.

In channels for transmission of television, noise is measured separately from periodic interference, primarily from the AC background, since this background is normalized separately. To do this, when noise is measured it is recommended that the high frequency filter used have a cutoff of 10 KHz. The suppression of low frequency noise which thus occurs has little influence on the result due to the great bandwidth of the video spectrum.

Finally, we note that measurement of noise is performed using special equipment when a video signal is present, since the noise with a modulated HF signal is somewhat different from the noise when there is no modulation. In practice, however, frequently noise is calculated and measured only with no modulation. Recommendation 421 of the ICCR permits this.

The ratio of signal to periodic noise from power supplies (frequency 50-100 Hz) should be at least 30 db[1]; in this case, in contrast to fluctuating noise, the ratio of signal amplitude to noise amplitude is normalized. The recommendation is related to the background which is added tothe signal (but does not modulate it) and can be eliminated by a levelcoupling device, measurements being performed with the damper disconnected. This recommendation influences the carefulness of development and adjustment of apparatus, but is not very significant in the selection of the primary parameters of the radio line. On the other hand, the fulfillment of the recommendation as to background level is very important for the transmitting stations of satellite systems, since these lines generally are quite long and connect points with nonsynchronized power supplies.

The ratio of the signal to periodic noise in the frequency band from 1 KHz tol MHz should be at least 50 db.

Nonlinear distortions of the image signal during a line are measured using a two-component test signal: a saw-toothed signal with an amplitude equal to the image signal (not including synchronization pulses) and a length approximately equal to the length of a line, on which a test high frequency signal of small amplitude /149

[1]For terrestrial systems, according to Soviet standards, the requirement is from 35 db for lines 2500 km long to 30 db for lines 12,500 km long.

(0.1 v) is superimposed, called the filler[1]. The frequency of the filler in a case of black and white television is 20% of the highest image frequency F_B, i.e., 1.2 MHz for the 6 MHz video band which we use. If there are nonlinear distortions in the line, the transmission factor depends on the instantaneous value of the image signal and modulation of the filler occurs. A filter is used to separate the filler from the saw-toothed signal, and the minimum m and maximum M values of filler amplitude are estimated from the screen of an oscilloscope. The nonlinear distortions are calculated in percent using the formula

$$(1 - \frac{m}{M}) \cdot 100\% \qquad (6.1)$$

and should not exceed 20% for the entire hypothetical circuit. This corresponds to the requirements for surface lines over 2500 km in length.

Recommendation 421 of the ICCR also normalizes changes in synchronization signal amplitude resulting from nonlinearity. This change should not cause the amplitude of the synchronization signals to go beyond limits of 0.21-0.33 v when the nominal value is 0.3 v (i.e., the deviation should not exceed +10% or -30%). In satellite communications systems, this requirement should be fulfilled, but generally places no additional conditions on the line itself, since the terminal receiver stations, which are complex and expensive structures, almost always are able to regenerate the synchronization signal; in some systems, the standard synchronization mixture is not transmitted through the line at all, in order to improve the power characteristics of the line for transmission of the accompanying sound or other purposes.

The linear distortions of the signal in the stable mode are normalized as follows. Distortions of the amplitude-frequency characteristic[2] should not exceed ±1 db in the 50 Hz-1 MHz band, ±2.2 db at 5 MHz, +4 -2.5 db at 6 MHz (see Figure 6.2). For surface lines 2500 km long, some increase in the distortion at 5 MHz and 6 MHz is permitted (by 1.1 db and 2 db). Distortions of group delay time characteristics should not exceed the limits shown in Figure 6.3. Furthermore, the reaction of the line to transient effects is tested and considered very important. Distortions of the flat top of a pulse with a length one half the field length are tested (they should be not over ±5%), as well as the nature of short-term distortions of the leading edge of a pulse with a rise time $T = 1/2F_B$ and an amplitude equal to the /150

[1]Damping and synchronizing pulses are usually also transmitted.

[2]Having in mind the dependence of attenuation on frequency.

signal amplitude (variations of not over +20% are required, while the length of the leading edge of the pulse at the output of the system should not exceed 0.14 μsec). Recommendation 421 of the ICCR establishes standards for evaluation of permissible distortions (see also [6.1]).

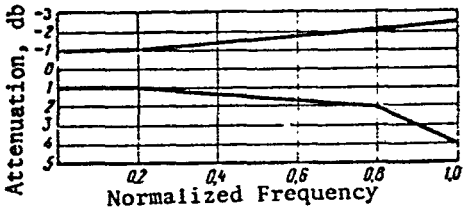


Figure 6.2. Limits on Amplitude-Frequency Characteristic for Television Systems

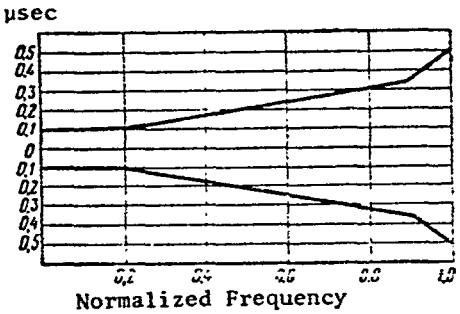


Figure 6.3. Limits of Characteristics of Group Delay Time of Television Systems

At the input and output of the entire line, as well as points of connection of various sets of apparatus, it is desirable to observe the value of full video signal amplitude of 1 v recommended by the ICCR, the amplitude of the image signal being 0.7 v, of the synchronization signal -- 0.3 v. The residual attenuation of the entire line should be 0 ± 1 db, brief changes should not exceed 0.3 db.

The performance requirements of the television sound channel have not been normalized for satellite communications lines on the international scale. Recommendation 402 of the ICCR merely assumes that the sound channel of television signals in radio relay lines corresponds to the IT-TCC requirements for international broadcast channels. The norms effective in the USSR and GOST 11, 515-65 for radio broadcast lines also assume the same quality of sound signal in long-range television transmission lines as in international first class broadcast exchange channels. /151

The band of frequencies of such channels should extend from 50 through 10,000 Hz with unevenness (as to attenuation) of +0.5 - -0.2 nep in the 50-100 Hz and 8.5-10 KHz bands; +0.3 - -0.2 nep in the 100-200 Hz and 6.0-8.0 KHz bands; and ±0.2 nep in the 200 Hz-6 KHz band. If the channel includes a predistorted frequency characteristic, it should be included along with the restoration circuit when unevenness of the amplitude-frequency characteristic is checked; the control circuits of dynamic range compressors and expanders should be blocked.

The signal-to-noise ratio in a first class international broadcast channel should be 6.55 nep (57 db) when measurements are made through a psophometric filter as recommended by the ICCR for a radio broadcast channel. The domestic standard (GOST 11,515-65) defines a value of 63 db without a psophometric filter as the norm. These two values approximately correspond with a triangular noise spectrum, since the broadcast channel psophometric filter has a rise at high frequencies (transmission factor greater than 1) and the signal-to-noise ratio decreases in psophometric measurements.

It should be kept in mind that the suppression of noise created by compander systems, broadly used in radio broadcast channels, is particularly great during a pause in transmission, i.e., under conditions corresponding to those used in measurement of noise. The actual suppression of noise during operation is less. Keeping this in mind, it is recommended that when noise is measured in channels with companders during a pause in transmission, the results produced be compared with a norm, increased by 1.35 nep (i.e., 6.55 + 1.35 = 7.90 nep).

These norms relate to the sum of thermal and unintelligible transient noises. Intelligible transient noises are significantly more noticeable; therefore, the norm set for these noises is more rigid -- 8.5 nep (74 db) without a weighting filter.

The relative background noise level (measured without the psophometric filter) should be not over -55 db; the norm can be dropped to -52 db for lines over 2500 km long.

In the case of transmission of color television signals, certain additional requirements must be placed on the channel, in spite of the compatibility of color and black and white television and the attempts made by color television system engineers to exclude these requirements. Requirements for color television transmission channels have not yet been established internationally.

The transmission band -- up to 6 MHz -- has already been selected for our purposes considering the transmission of color television, since band limitations in this case result in losses not only of clarity but in color quality as well. Therefore, all /152
requirements established earlier for the transmission band and linear distortion should be unwaveringly adhered to in the transmission of color television signals.

The requirements as to residual attenuation of the line and its stability, level of periodic interference, etc. should not be changed.

Fluctuating noise is more strongly observed in a color picture than in a black and white picture. Nevertheless, the norms

for signal to fluctuating noise ratio established earlier (55 db) are high enough that the quality of the color image will remain acceptable. A norm of 57 db, as established by Recommendation 421 of ICCR, is more desirable, particularly when color transmission is by the SECAM-3 system used in the USSR, which has relatively low interference stability[1]. Measurement of weighted noise can be performed using a standard weighting filter although, strictly speaking, a specialized filter would be more suitable for color television [6.1]. In particular, it is desirable to test the signal-to-noise ratio in the frequency band of the color signal.

Nonlinear distortions of the picture signal are particularly important for color television. Actually, a change in the high frequency filler as a function of the instantaneous value of the low frequency component in this case means a change in the level of the color subcarrier (and, consequently, the color saturation) as a function of the brightness signal. Therefore, measurements are performed at a filler frequency not of 1.2 MHz as in the case of black and white television, but rather at the frequency of the color subcarrier (4.4 MHz). This in itself means an increase in the requirements, since distortions generally increase with modulation frequency, and when measurements are performed by a two-frequency method -- with filler frequency [6.2]. Furthermore, the norm itself is usually more rigid, dropping from the 20% indicated earlier to 5-10%. In the case of color television, a special term -- differential gain -- following from the essence of the two-component method of measurement, is used in place of the term "nonlinear distortion."

Another specific channel performance factor for color television which is quite similar in its significance to differential gain is differential phase, defined on the basis of the same two-component test signal as the maximum filler phase change (at 4.4 MHz) during the period of the saw-toothed signal. Differential phase distortions cause changes in hue and are therefore quite dangerous. The SECAM system is relatively insensitive to these distortions, but it is still desirable that the differential phase not exceed $\pm 10^{\circ}$[2]. As will be shown below, this condition /153 places particularly serious requirements on the high frequency section of FM satellite communications lines.

It is considered necessary to control the group delay of the color signal relative to the luminance signal. This delay should not exceed 25-30 nanosecond.

---

[1]Other, more interference-stable color television systems may be used in the satellite line itself.

[2]And in the NTSC system, particularly sensitive to differential phase, $\pm 2^{\circ}$ [6.3, 6.4].

Finally, it should be kept in mind that the amplitude of a color video signal, due to the imposition of the brightness and color signals, is somewhat greater than that established for a black and white signal (0.7 v), amounting 0.86 v in the SECAM system, causing the total amplitude of the television signal including synchronization pulses to reach 1.16 v.

§ 6.3. Performance Factors of Satellite Communications Channels for Transmission of Telephony

The band of frequencies of a telephone channel -- 300-3400 Hz -- is a firm international standard and cannot be changed. The attenuation at the extreme ends of the frequency band should not exceed the attenuation at 800 Hz by more than 1 nep.

The most important factors is the permissible noise level. Recommendation 353 of the ICCR[1] calls for a psophometrically weighted noise power at the point with 0 relative level (0 db, 1 mw) for any telephone channel of the primary hypothetical standard circuit in a satellite communications system of not over 10,000 pw, averaged for any hour. Averaged per minute, noise must not exceed 10,000 pw for more than 20% of any month, 50,000 pw for more than 0.3% of any month (i.e., for 0.3% or less of the time, noise may exceed 50,000 pw); unweighted noise must not exceed 1 million pw for more than 0.03% of any month (integration time 5 msec).

The noise level in satellite communications lines is generally quite stable; therefore, primary calculations should be performed for the hourly mean norm of 10,000 pw. On the other hand, it should be kept in mind that unfavorable propagation conditions (strong rain, for example), which amount to a very small fraction of the time over the long run, may increase the hourly mean noise level.

Therefore, the ICCR norms are not relaxed for satellite lines; Recommendation 394 calls for the same hourly mean noise power for a hypothetical standard radio relay circuit 2500 km long, while the Soviet standards for radio relay and cable main lines yield similar values (8100 pw). We might consider that the norms are slightly relaxed by the fact that the 10,000 pw norm for satellite lines does not include multiplexing apparatus noise. Considering that satellite lines usually cover distances much longer than 2500 km, we must assume that fulfillment of the noise norm on these lines will actually yield an even higher quality signal than when surface lines are used. It is therefore considered possible to use a noise norm of 20,000 pw for satellite communications lines. The noise in the communications channel results not only from thermal noise in the input circuits /154

[1] Revised in 1966 at the ICCR Conference in Oslo.

of the receivers (both on board the satellite and at ground stations), but also crosstalk interference between channels. Thus, in the case of multichannel telephony with frequency-division multiplexing, crosstalk interference results from nonlinearities in the group line. The norm indicated above must include the sum of thermal and nonlinear noises.

Selection of the optimal relationship between these two types of noise in a channel is one of the basic tasks in planning of communications lines and will be discussed separately in Chapter 9.

Furthermore, since common frequency ranges are used for satellite and radio relay lines, mutual interference between them may occur. The ICCR recommends[1] that interference caused by RRL produce noise with a psophometric mean power during any hour of not over 1,000 pw in any telephone channel of a hypothetical standard satellite communications line.

As in the case of a radio broadcast, norms for intelligible interference are still more rigid. For 90% of all channel combinations (interfering and interfered), this interference should be at most -6.7 nep. In multichannel communications systems with frequency-division multiplexing, intelligible noise in the satellite line generally cannot arise; however, in communications systems with multistation access with transmission of station signals on separate carriers, intelligible interference may occur.

The recommendation for terrestrial communications channels also define the nonlinear distortions permissible in the channel. Nonlinear distortions for telephone conversations themselves are not too dangerous, since they have little influence on intelligibility, while nonlinearities of the microphone and speaker in the television apparatus of a subscriber are quite great. However, with secondary multiplexing of a telephone channel (for example, of a large number of telegraph signals), nonlinearity can cause crosstalk interference and therfore, according to the Soviet standards, should not exceed 1.5% (not over 1% in the third harmonic) over one sector of transmission and reception. A standard satellite line can be considered equal to two or three sectors of transmission and reception of a surface line, and the norm increases with the number of sectors n according to $1.5\sqrt{n}$%.

The ITTCC merely provides an indirect estimate of the linearity of a telephone channel, based on its amplitude characteristic. The net attenuation in the channel should remain constant with an accuracy of 0.035 nep as the level at the input of the channel changes from -2.0 to +0.4 nep (at the zero relative level point. In order to prevent overloading of multichannel systems /155

[1]See Chapter 10.

with individual channels, it is recommended that a limiter be set at the input of each long-range (four-wire) telephone channel, increasing the net attenuation in those cases when the input level exceeds +0.4 nep. This measure is also quite useful for satellite communications systems with multistation access, using time or frequency separation of subscribers' signals, including (even particularly) with transmission of one channel in each direction.

During specific planning of the equipment for communications lines, the nominal values of levels at various points in the channels are significant, including the measured transmission level at the two-wire input of the channel (0 nep), upon transmission into the four-wire portion of the channel -1.5 nep, upon reception +0.5 nep; at transit switching points in the channel -0.4 nep.

The instability of relative attenuation with time, measured at 800 Hz, should not exceed ±0.12 nep for each transmission-reception sector (mean-square deviation).

One specific feature of satellite lines is the long group propagation time. In a single-hop line (i.e., one Earth-satellite-Earth step) with a stationary satellite or a Molniya-1 satellite (at apogee), this time reaches 300 msec. This amount of time is permitted by the recommendations, although contact between speakers is made more difficult and requires some skill. Echo suppressors must be introduced to the channel in this case, since otherwise the delayed echo makes conversation quite impossible.

A delay of over 500 msec is not permitted by Soviet recommendations, i.e., lines with two hops using high-altitude satellites cannot be used for telephone communications.

The norms limiting the radiated power of satellite and ground station transmitters in order to avoid excessive interference with radio relay communications lines are very significant in planning satellite systems. These recommendations are presented in Chapter 10.

REFERENCES

6.1. Krivosheyev, M. I., Osnovy televizionnykh izmereniy [Principles of Television Measurement], Svyaz' Press, 1964.

6.2. Kantor, L. Ya., Metody povysheniya pomekhozashchishchennosti priyema ChM signalov [Methods of Increasing the Interference Stability of Reception of FM Signals], Svyaz' Press, 1967.

6.3. ICCR Report No. 316, Presented at 10th Plenary Assembly of ICCR.

6.4. Document XI/67-E, ICCR, 26 March 1965.

CHAPTER 7. THE TRANSMISSION POWER CHARACTERISTICS OF SATELLITE COMMUNICATIONS LINKS

§ 7.1. General Electrical Characteristics of Satellite Communications Lines

/156

Satellite communications links generally consist of two sections: the Earth-satellite and satellite-Earth sections the satellite-Earth section being energetically more constrained, since the power of the satellite transmitter is rigidly limited. Although the range of transmitter powers used on satellites is rather broad (depending on the purpose of the satllite and the nature of the signal, it extends from a fraction of a watt to several kilowatts), nevertheless in each case without exception decreasing this power to allow a reduction in satellite weight and size is of primary significance.

Another peculiarity of the lines in question is the presence of great losses of signal power over the propagation path, resulting from the great physical length of this path. For example, with a satellite orbital altitude of 40,000 km, the maximum distance between transmitter and receiver is 45,000 km, and the attenuation of the signal may reach 200 db.

In addition to attenuation in space, the signal in satellite communications lines is subjected to the influence of a very large number of other factors, such as absorption in the atmosphere, Faraday rotation of the plane of polarization, refraction, etc. On the other hand, the receiver device used for detection and separation of the signal is influenced by fluctuating noise plus various types of interference such as cosmic, solar, planetary and atmospheric radiation.

Under these conditions, proper and precise consideration of the influence of all factors allows us to plan systems optimally to assure reliable operation under the most difficult conditions while at the same time eliminating unnecessary power reserves, which lead to unjustified increases in the complexity of ground station and satellite equipment.

/157

Furthermore, as was noted in Chapter 6, the recommendations for the performance factors of satellite communications channels are statistical in nature. This forces us to consider all possible perturbing factors statistically and to use the quantitative measure of the influence of a factor as well as its probability (frequency) of appearance in our calculations.

Finally, we must consider the nature of the conversion of the signal on board the satellite. In the simplest case, no signal processing is performed on board and the satellite is simply a passive relay station. Communications lines with passive relay are under more difficult power conditions, since a passive relay station reradiates only a portion of the power of the ground transmitter, while the losses of energy in space correspond to two link sections.

The energy potential of a line is significantly improved if simple processing of the signal -- amplification -- is used on board the satellite. However, this active reradiation is not the optimal form of processing, since it is accompanied by the accumulation of noise over two sectors of the communications line.

The best conditions, obviously, are provided by full regeneration of the signal on the satellite; the expediency of this processing is determined in each concrete case by the permissible degree of complication of the on-board apparatus.

This peculiarity of relay prevents calculations using the same formulas and requires analysis of each system individually.

However, all three types of relay share certain general features, such as the principle of performance of communications, propagation conditions of radio waves and, finally, identity of block diagrams, as well as, at times, the parameters of the ground station equipment. These facts allow us to design lines according to a single methodology, basing the calculation on one sector of the communications line.

§ 7.2. Transmission Equation for One Section

A block diagram of a section of a communications line and a diagram of the signal levels are presented in Figure 7.1. The transmitter develops effective power P_{tr} at frequency f (wavelength λ) at its output. This signal, passing through the antenna-waveguide system, with power transmission factor η_{tr}, arrives at the antenna and is radiated into space. The capability of the transmitting antenna to concentrate the energy transmitted in a given direction is characterized by its gain G_{tr} in comparison to an undirected (isotropic) radiator. Thus, the effective power of the signal radiated by the transmitting antenna in the required direction is

$$P_{rad} = P_{tr}\eta_{tr}G_{tr}.$$

The effective electromagnetic field intensity E_0 at distance d from the transmitter with propagation in free space /158

$$E_0 = \frac{\sqrt{30P_{rad}}}{d}, \qquad (7.1)$$

while the attenuation of energy in free space, determined by the decrease in radiation flux density with increasing distance from the radiator

$$L_0 = \frac{16\pi^2 d^2}{\lambda^2}. \qquad (7.2)$$

In addition to these primary losses, there are unavoidably additional losses along the propagation path, resulting from attenuation of the energy in the atmosphere, as well as the influence of various perturbing factors. These losses will be represented by L_{add}. Thus, the total power losses $L_\Sigma = L_0 L_{add}$.

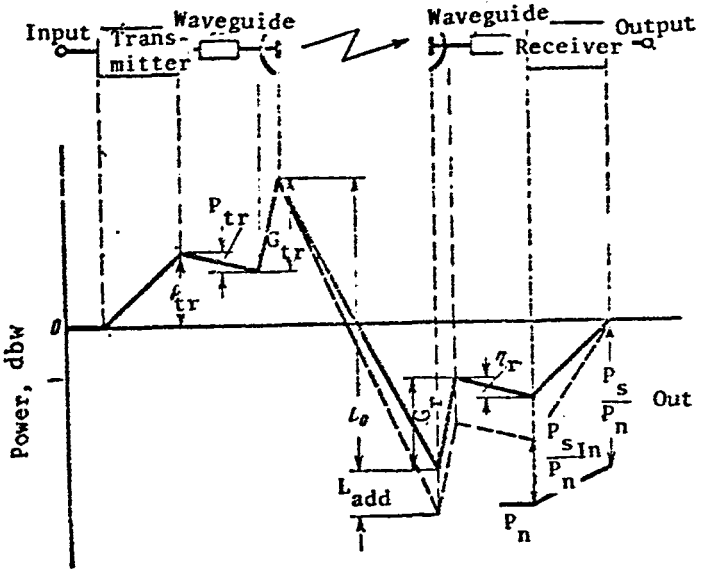


Figure 7.1. Block Diagram and Level Diagram of One Sector of a Communications Line

At the reception point, there is an antenna with gain G_r, connected to the receiver by an antenna-waveguide system with power transmission factor η_r. With proper matching of the wave impedances /159 of elements of the system, the signal power at the input of the receiver will be

$$P_r = P_{tr} \frac{\lambda^2 G_{tr} G_r \eta_{tr} \eta_r}{16\pi^2 d^2 L_{add}}.$$

This form of this expression is the generally known formula used for calculation of line of sight radio lines.

Continuing the development of this formula, we can state that in the case when the parameters of the receiving antenna are fixed not in the form of the gain, but rather in the form of the effective area S_e, which is related to the gain by the formula

$$G_r = \frac{4\pi S_e}{\lambda^2},$$

the preceding expression can be written as

$$P_{tr} = P_r \frac{4\pi d^2 L_{add}}{G_{tr} S_e \eta_{tr} \eta_r}. \qquad (7.3)$$

Formula (7.3) allows us to determine the necessary transmitter power on the basis of the required value of signal power at the input of the receiver.

It is interesting to note that (7.3) does not include the wavelength. Consequently, when the transmitter antenna has a constant gain at all frequencies, while the receiving antenna has a constant effective area (constant geometric dimensions) and can be used with beams ever more constricted with increasing frequency, the signal power at the input of the receiver is independent of frequency in the first approximation, i.e.,

$$P_r = \text{const} \frac{P_{tr}}{d^2}.$$

Actually, some frequency dependence still occurs, since the additional losses L_{add} are determined to a significant extent by the frequency range.

It is sometimes not the minimum signal power, but rather the minimum permissible signal/noise ratio at the input of the receiver $(P_s/P_n)_{in}$ which is fixed. In this case, we must replace P_r in formula (7.3) with its value from the relationship

$$P_r = P_n(P_s/P_n)_{in},$$

where P_n is the power of noise at the input of the receiver.

The noise power is the quantity which, essentially, limits the sensitivity of receiver systems and defines the minimum level of the signal in the communications line.

A detailed quantitative estimate of all noise components in the satellite communications receiver will be presented below. We can state preliminarily that, since the noise created by the various sources in the range of frequencies in question is additive in nature, the summary noise power should be expressed by the formula /160

$$P_n = kT_\Sigma \Delta f_n, \qquad (7.4)$$

where $k = 1.38 \cdot 10^{-23}$ w/Hz·deg is Boltzman's constant;
T_Σ is the equivalent noise temperature of the entire receiver system (considering internal and external noises);
Δf_n is the equivalent frequency band of noise in the receiver.

Formulas (7.3) and (7.4) establish the relationship between the primary parameters of the line and are the initial relationships used in planning.

For a rapid and approximate estimation of the parameters of a line, it is convenient to use the nomogram of Figure 7.2.

§ 7.3. Transmission Equation with Active Relay

As we have already stated, with active relay the communications line consists of two sections coupled by the relay station, which acts as a signal power amplifier (Figure 7.3). Using the formulas

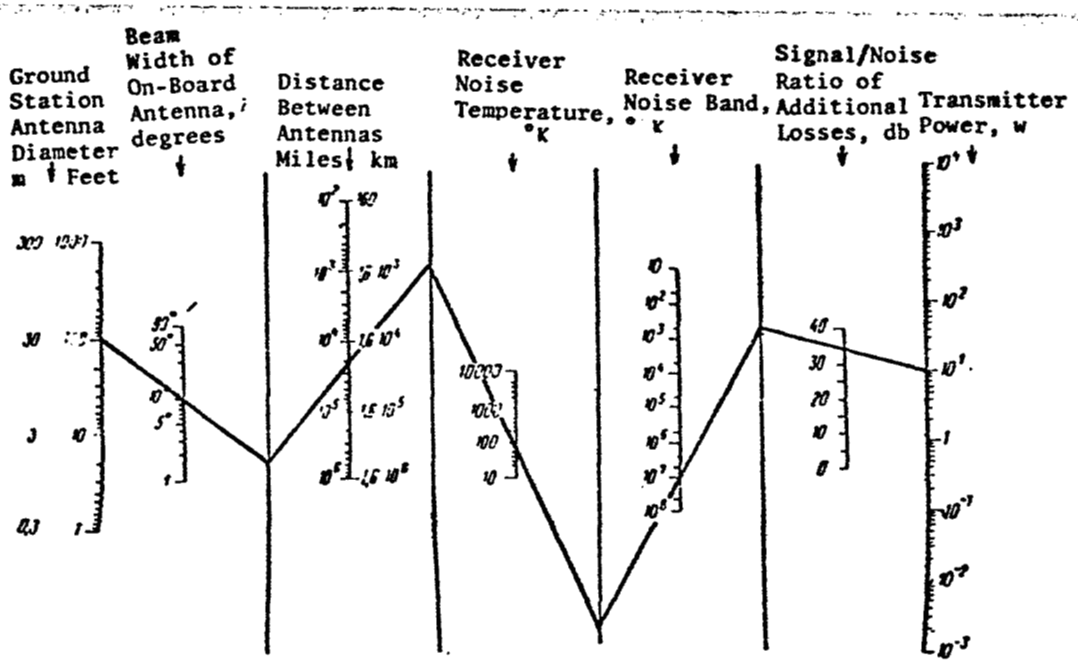


Figure 7.2. Nomogram for Design of Satellite Communications Lines (Receiver Noise Band Given in Hz)

presented in § 7.2, we can write the following relationships:

-- for the Earth-satellite sector[1]

$$P_{trg} = \frac{16\pi^2 d_1^2 L_{1add} P_{ns}}{\lambda_1^2 G_{trg} G_{rs} \eta_{trg} \eta_{rs}} (P_c/P_n)_{ins},$$

where $P_{ns} = kT_s \Delta f_{ng}$;

-- for the satellite-Earth sector

$$P_{trs} = \frac{16\pi^2 d_2^2 L_{2add} P_{ng}}{\lambda_2^2 G_{trs} G_{rg} \eta_{trs} \eta_{rg}} \left(\frac{P_s}{P_n}\right)_{ing},$$

where $P_{ng} = kT_{ng} \Delta f_{ng}$.

[1]Here and below, all characteristics relating to the ground apparatus will be marked with a subscript "g," characteristics relating to the satellite apparatus will be marked with the subscript "s."

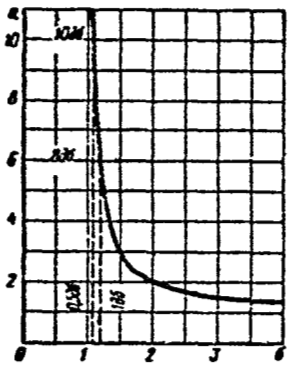


Figure 7.4. Distribution of Signal/Noise Ratio Over Two Sections of the Communications Link

In order to go over from equations for individual sectors to the overall equation for the entire line, we must establish the relationship between the signal/noise ratios at the input of the satellite receiver, the ground receiver and at the end of the entire communications line. As was noted above, in a satellite line without regeneration of the signal on board the satellite, the noises present in each sector are added; therefore, for the sum of both sectors we can write

$$\left(\frac{P_n}{P_s}\right)_\Sigma = \left(\frac{P_n}{P_s}\right)_{ins} + \left(\frac{P_n}{P_s}\right)_{ing} \qquad (7.5)$$

Obviously, the signal/noise ratio in each sector should be higher than at the end of the communications line. Let us introduce noise reserve factors a and b, defined as follows /163

$$\left(\frac{P_s}{P_n}\right)_{ins} = a\left(\frac{P_s}{P_n}\right)_\Sigma$$

$$\left(\frac{P_s}{P_n}\right)_{ing} = b\left(\frac{P_s}{P_n}\right)_\Sigma. \qquad (7.6)$$

Substituting (7.6) into (7.5), we can find the relationship between a and b. The corresponding dependences shown in Figure 7.4, using which we can determine the summary signal/noise ratio required for the two sectors of the communications line. In practice, b should be selected as 1.1-1.25 (0.5-1 db), allowing a to be 10-6 (10-8 db); with any different values, the transmitter required at the ground station or on the satellite will be unjustifiably large.

The distribution of reserve factors a and b presented assumes that the frequency bands of the satellite and ground receivers are identical. In practice, the band of frequencies of the relay station is always broader than that of the ground receiver. In this

case, the noise power at the input of the satellite receiver should be defined in the band of frequencies of the ground receiver.

/162

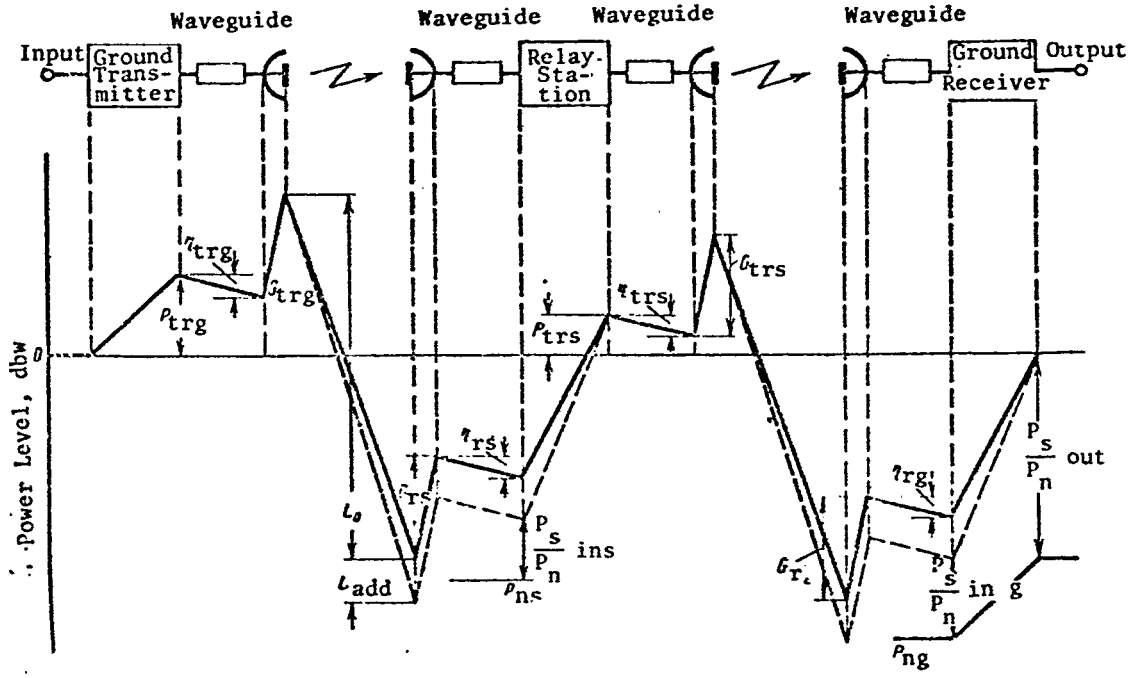


Figure 7.3. Block Diagram and Diagram of Levels of Line with Active Relay

Considering what we have stated, the coupling equation for lines with active relay can be finally written as:

-- for the ground-satellite sector

$$P_{trg} = \frac{16\pi^2 d_1^2 L_{1add} kT_s \Delta f_{ng}}{\lambda_1^2 G_{trg} G_{rs} \eta_{trg} \eta_{rs}} a\left(\frac{P_s}{P_n}\right)_\Sigma, \qquad (7.7)$$

-- for the satellite-ground sector

$$P_{trs} = \frac{16\pi^2 d_2^2 L_{2add} kT_g \Delta f_{n.g}}{\lambda_2^2 G_{trs} G_{rg} \eta_{trs} \eta_{rg}} b\left(\frac{P_s}{P_n}\right)_\Sigma. \qquad (7.8)$$

For an approximate estimation of the parameters of the line, we can use the nomogram of Figure 7.2 as before, using it to calculate the parameters of each sector and then introducing corrections resulting from the redistribution of noises in the two sectors of the communications line. /164

All formulas produced in this section are also suitable for lines with regeneration of the signal on the satellite. In this case, we should assume a = b = 1, since regeneration, as was stated above, eliminates the accumulation of errors.

§ 7.4. Transmission Equation with Passive Relay

With passive relay of signals, the power of the transmitter radiated toward the satellite is reflected from its surface according to the laws of geometric optics. A portion of this energy strikes the receiving antenna on the surface of the Earth and forms the signal at the input of the receiver.

In order to analyze the processes occurring during passive relay, let us assume that the diameter of the passive satellite reflector D_0 is significantly greater than wavelength λ, while the incident wave, due to the significant distance between the radiator and reflector, can be considered flat, with constant amplitude and phase.

The power flux density striking the reflecting surface, separated from the transmitting antenna by distance d_1, is equal to

$$W_{tr} = \frac{P_{tr}\eta_{tr}G_{tr}}{4\pi d_1^2 L_{1add}}. \qquad (7.9)$$

If the width of the main lobe of the radiation pattern of the transmitting antenna is greater than the angle at which the object is visible from the Earth, the incident energy irradiates its entire reflecting surface. We note that this is almost always the case in practice.

The density of reflected energy W_r will depend both on the reflection factor κ and on the effective reflecting surface of the relay satellite. This surface is determined by its shape and size. Without discussing problems of selecting the optimal design and shape for a reflector, let us introduce parameter σ, characterizing this effective reflecting surface.

The density of the reflected energy at distance d_2 from the relay satellite, by analogy with (7.9), will be

$$W_r = \frac{W_{tr}\sigma\kappa}{4\pi d_2^2 L_{2add}} = \frac{P_{tr}G_{tr}\sigma\kappa}{16\pi^2 d_1^2 d_2^2 L_{1add}L_{2add}}.$$

If the effective area of the receiving antenna is S_r, the signal power at the input of the receiver

$$P_r = \frac{P_{tr}G_{tr}\sigma S_r \kappa \eta_{tr}\eta_r}{16\pi^2 d_1^2 d_2^2 L_{1add}L_{2add}} \qquad (7.10)$$

and where $d_1 = d_2$ it is inversely proportional to the fourth power /165
of the distance between the ground station and the satellite.

It is interesting to note that if the dimensions of the relay satellite are so great that the transmitting antenna illuminates only a portion of its reflecting surface, the energy loss in the Earth-satellite sector will be zero[1], since all of the energy concentrated in the main lobe of the radiation pattern is reradiated by the satellite. The power of the signal at the input of the receiver in this case is inversely proportional only to the square of the distance between the satellite and the ground station antenna.

Actually, substituting in (7.10)

$$\sigma = \frac{\pi D_0^2}{4} \approx \frac{\pi d_1^2 \varphi_{0.5}^2}{4},$$

where $\phi_{0.5}$ is the width of the radiation pattern transmitted by the ground antenna at the half power level, and considering that $G = \frac{\text{const}}{\phi_{0.5}^2}$, we produce /166

$$P_r = \frac{P_{tr}S_r \kappa \eta_{tr}\eta_r}{64\pi d_2^2 L_{1add}L_{2add}}.$$

[1] With the exception of the additional losses.

Returning to formula (7.10) and replacing P_r with the noise power at the input of the receiver and the required signal/noise ratio, we produce finally

$$P_{tr} = \frac{16\pi^2 d^4 L_{add}^2 kT_g \Delta f_{ng}}{G_{tr} \sigma S_r \kappa \eta_{tr} \eta_r} \left(\frac{P_s}{P_n}\right)_{in}. \qquad (7.11)$$

A block diagram of a communications line with a passive relay satellite and the associated level diagram are presented in Figure 7.5. Comparison of Figure 7.5 with Figure 7.3 clearly shows the power advantages of active relay.

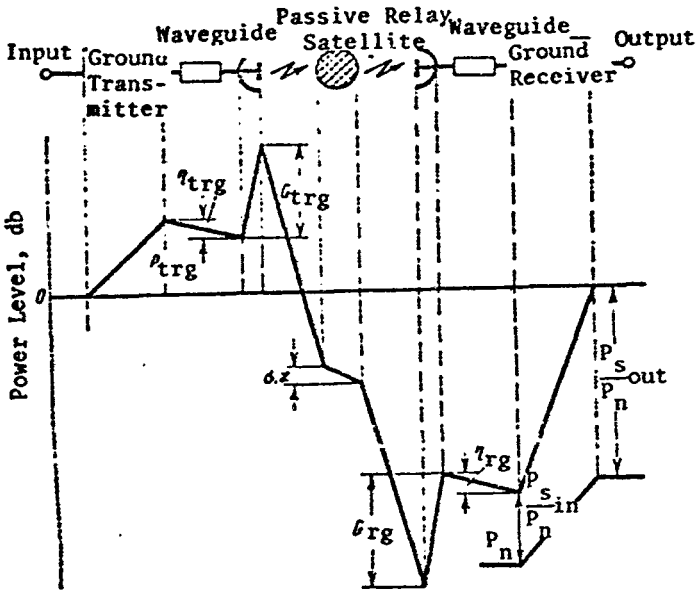


Figure 7.5. Block Diagram and Diagram of Levels of Communications Line with Passive Reflection

The formulas presented allow us to perform power calculation for satellite communications lines with various types of relay of signals. In order to produce quantitative results, let us study the basic quantities included in this formula.

§ 7.5. Calculation of the Gain of Ground Station Antennas

The technical requirements placed on transmitting antennas for ground stations can be reduced to production of the maximum gain. Receiver antennas, in addition to the maximum gain (or effective area), must also have good spatial selectivity and minimum level of side lobes.

These requirements are fulfilled differently, depending on the types of antennas used. There are several types of SHF antennas, but the commonest types in satellite radio communications systems are parabolic, horn-parabolic and double-mirror types[1].

The gain of an antenna is related to its geometric dimensions by the relationship

$$G = q\frac{4\pi S_A}{\lambda^2}\eta_A,$$

where S_A is the area of the aperture of the antenna, m^2;

q is the surface utilization factor;

η_A is the efficiency of the antenna (usually near 1).

Expressing the area of the antenna through its diameter /167

$$S_A = \frac{\pi D_A^2}{4},$$

we produce

$$G = q\frac{\pi^2 D_A^2}{\lambda^2}\eta_A. \qquad (7.12)$$

The surface utilization factor of an antenna depends on the type and design of the antenna. For well-designed parabolic antennas, this quantity amounts to 0.5-0.7; in horn-parabolic antennas, q = 0.7-0.8 can be achieved. The width of the radiation pattern of the antenna at the half power level is determined by the approximate formula

$$\varphi_{0.5} = \frac{70\lambda}{D_A} \qquad (7.13)$$

[1] Detailed information concerning types and design peculiarities of antennas is presented in the second volume of this work.

or, considering (7.12)

$$\varphi_{0.5} \approx \sqrt{q \frac{49\,600}{G}}. \qquad (7.14)$$

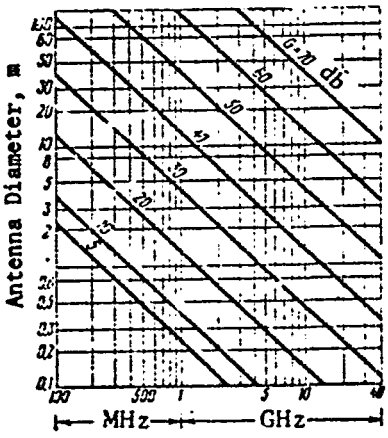


Figure 7.6. Gain as a Function of Frequency for Antennas of Various Diameters

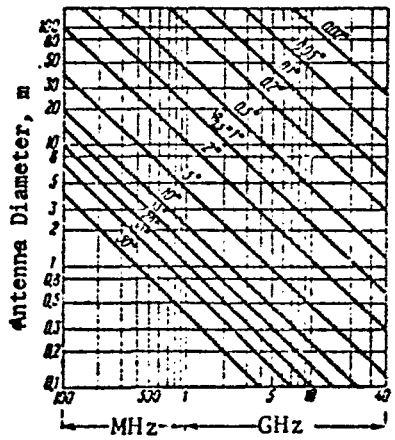


Figure 7.7. Radiation Pattern as a Function of Frequency for Antennas of Various Diameters

Figures 7.6 and 7.7 show nomograms constructed from formulas (7.12) and (7.13), allowing us to estimate the parameters of antennas of fixed diameters approximately.

We note that formulas (7.13) and (7.14) were produced on the assumption that the radiation pattern of the antenna was the same within the limits of the main lobe in both planes. /168

In the general case

$$\phi_{0.5} = \sqrt{\phi_{0.5E}\phi_{0.5H}},$$

where $\phi_{0.5E}$ is the width of the main lobe in the plane of the electrical field,

$\phi_{0.5H}$ is the width of the main lobe in the plane of the magnetic field.

The values of gain indicated above are fully realized only if the transmitting (or receiving) antenna is oriented precisely toward the satellite. Actually, as the satellite is tracked, guidance errors arise unavoidably, resulting in loss of gain. In order to estimate these losses quantitatively, let us study the radiation pattern of the antenna. A simplified expression describing the shape of the main lobe of the antenna is [7.1]

$$G(\varphi) \approx 2\frac{J_1(kR\sin\varphi)}{kR\sin\varphi}, \qquad (7.15)$$

where R is the radius of the aperture of the antenna,
$k = 2\pi/\lambda$ is the wave number,
$j_1(x)$ is a Bessel function of the first kind of the real argument.

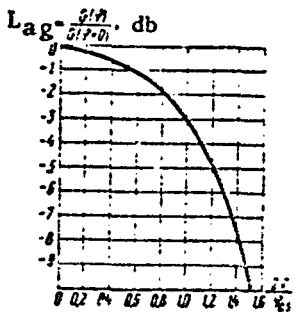


Figure 7.8. Form of Main Lobe of Radiation Pattern of Antenna

The results of calculations using this formula are presented in Figure 7.8, in which the vertical axis shows the ratio of the gain of the antenna in direction ϕ to the gain in the main direction $\varphi = 0$, while the horizontal axis shows the ratio of angle φ to the half width of the radiation pattern of the antenna at the half power level.

Using Figure 7.8, we can establish the relationship between the permissible drop in gain of the ground station antenna L_{Ag} and the angular error of its guidance $\Delta\phi_H$. For example, using loss value $L_{Ag} = -1$ db, we can easily find that $\Delta\phi_H \leq 0.3\ \phi_{0.5}$. If, however, the maximum error in guidance of the antenna is known, we can estimate the maximum value of L_{Ag}, which is a portion of the additional losses L_{add}.

We should note that the value of L_{Ag} does not remain constant during the process of operation of the system, since the error of orientation of the antenna with programmed guidance is related to its rate of rotation ω which, in turn, is determined by the rate of movement of the satellite relative to the ground stations. Although the true relationship between the angular rate of rotation of the antenna and the angular guidance error /169

is rather complex, the two quantities can be considered directly proportional in the first approximation.

This fact allows us to estimate the value of L_{Ag} statistically, i.e., to find its distribution function and to use this function to determine the maximum losses, which will not be exceeded for a fixed percentage of the time.

The angular velocity of a space object relative to the ground station

$$\omega(t) = |\bar{V}(t)\bar{r}| \frac{1}{d(t)},$$

where $\bar{V}(t)$ is the velocity of the object,
$\bar{r}$ is the unit radius vector directed toward the object,
d(t) is the distance to the object as a function of communications time (or session).

The method of calculation using this fomrula is not presented here, since these calculations are rather complex and cumbersome. We note only that in planning lines, the value of $\omega(t)$ can be easily determined from the ephemerides of the space object, calculated for each communications session.

As an example, Figure 7.9 shows the curves of the change of azimuth and elevation of an antenna located in the region of Moscow and oriented towards a Molniya-1 satellite. Using these dependences, we can easily construct curves of the change of angular velocity during the course of communications session T, then find graphically the integral distribution function of this velocity. The results of this statistical processing of the curves of Figure 7.9 are presented in Figure 7.10 (curve 3) in the form of an integral distribution function of the resulting angular velocity (resulting angular velocity is equal to geometric sum of azimuthal and elevation velocities). Here also we present curves of the distribution of velocities for antennas at Novosibirsk and Vladivostok (curves 2 and 1 respectively).

It follows from Figure 7.10 that the greatest angular velocity of movement of the Molniya-1 satellite during 99.9% of the time of a communications session is not over 0.2 deg/sec, while the mean velocity is approximately 0.004-0.01 deg/sec. Consequently, if the antenna orientation system has a maximum translation rate of, for example, ω_{max} = 1 deg/sec, providing a guidance error of not over 0.3 $\phi_{0.5}$ (as was established above, the power losses in this case amount to 1 db), for 99.9% of the time this error will not

/170

exceed 0.06 $\phi_{0.5}$, and the power losses, according to Figure 7.8, will be not over 0.1 db. The mean guidance error will be approximately 0.003 $\phi_{0.5}$ and the mean losses correspondingly will be almost equal to 0 db.

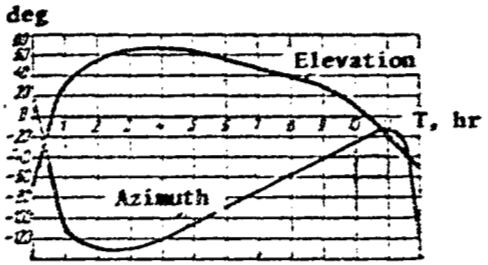


Figure 7.9. Change in Azimuth and Elevation During the Course of a Communications Session for an Antenna Located at Moscow and Tracking a Molniya-1 Type Satellite

It should be kept in mind that these discussions do not consider errors in determination of the ephemerides of the object and assume that the spatial position of the object is rather precisely known.

In the case when the guidance of the antenna is not programmed, but rather follows a beacon signal or the communications signal, these statistics are incorrect; however, the guidance error is rather slight in this case, approximately 0.01° in the best antennas.

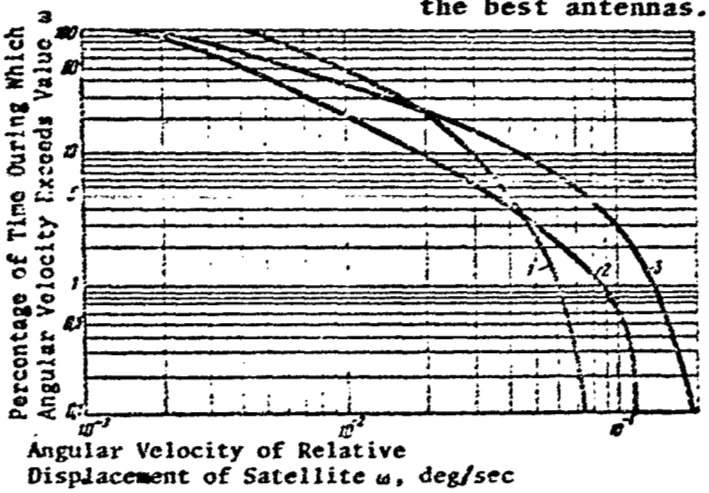


Figure 7.10. Distribution Function of Angular Velocity of Antenna Tracking Molniya-1 Type Satellite

C-3

§ 7.6. Calculation of Gain and Shape of Satellite Antenna Radiation Pattern

The same requirements are placed on satellite antennas as on ground station antennas: they include maximum concentration of the satellite transmitter power within a certain solid angle. However, in contrast to ground station antennas, the width and shape of the radiation pattern of the satellite antenna must be selected in accordance with the size and shape of the surface of the Earth to be covered. /171

The width of the radiation pattern (at the half power level) of a satellite antenna moving in a circular orbit at altitude H and oriented toward the center of the Earth with angular error γ can be determined according to the following formula, illustrated by Figure 7.11:

$$\varphi_s = 2 \arcsin \frac{R_g}{R_g + H} \cos \beta_{min} + 2\gamma, \qquad (7.16)$$

where R_g is the radius of the Earth,

β_{min} is the minimum elevation of the ground station antenna at the edge of the zone of coverage.

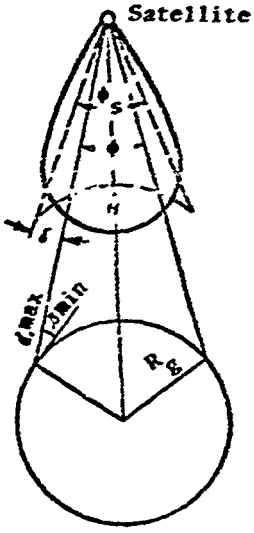


Figure 7.11. Calculation of Satellite Antenna Radiation Pattern

The gain corresponding to the width of the radiation pattern can be defined both by formula (7.14) and using the nomograms of Figures 7.6 and 7.7.

However, in order to calculate power requirements for the line, we must know not only the gain of the antenna in the primary direction, but also the unevenness of gain within the limits of the working portion of the main lobe of the radiation pattern.

According to the calculation method used with formula (7.16), the entire territory of the Earth covered by the satellite lies within the solid angle within which the gain of the antenna changes by a factor of not over 2. Consequently, the losses in satellite antenna gain L_s resulting from unevenness of the

radiation pattern with a fixed orientation error will not exceed 3 db.

In order to eliminate these losses, it is possible in principle to synthesize a radiation pattern with a special shape, assuring a constant satellite signal power flux within the limits of the entire zone serviced (this diagram is illustrated on Figure 7.11 by the dotted line). However, synthesis of this type of radiation pattern is rather difficult and has not been used to date.

For elliptical orbits, calculation of the gain and radiation pattern is more complex. This results from the fact that as a satellite moves through a noncircular orbit, both the altitude of the satellite and the visible area change. An additional complication is introduced by the fact that the error in antenna orientation γ also varies with satellite altitude. /172

For points at the shortest distance from the satellite (the point beneath the satellite), the signal power flux will change only due to changes in distance. However, for points at the edge of the zone of visibility, the signal power flux will change both due to changes in range, and due to changes in the solid angle in which the Earth is visible from the satellite, since with a fixed radiation pattern width, a change in the solid angle causes a change in the gain of the satellite antenna in the required direction.

As a quantitative illustration of this, let us calculate the change in signal power flux as a satellite moves through an orbit such as that used by the Molniya-1 satellites for points on the Earth's surface at the minimum and maximum distances from the satellite, using standard parabolic antennas with radiation patterns as described by expression (7.15).

Calculation of the radiation pattern width will be performed using formula (7.15), assuming that at the initial point of the orbit, the apogee, H = 40,000 km, $\beta_{min} = 7°$, while $\gamma = 3°$, we produce $\phi_s = 21.5°$, $G_s = 18$ db. The power of the signal at the input of the receiver, according (7.8), is equal to

$$P_{sin} = \frac{G_{trs}[\phi(t), \gamma]}{d^2(t)}. \qquad (7.17)$$

The values of the instantaneous parameters $d_{max}(t)$, $d_{min}(t)$ and $\phi(t)$, read off at time intervals $\Delta t = 1$ hr from the moment when the Molniya-1 satellite passes through the perigee, are presented in Table 7.1.

TABLE 7.1

t, hr	1	2	3	4	5	6
d_{min}, km	13 450	24 200	31 530	36 330	39 100	39 990
d_{max}, km	18 020	29 130	36 000	41 450	44 250	45 130
ϕ, deg	37° 16′	23° 54′	19° 06′	17° 08′	16° 04′	15° 42′

Using these data, it is not difficult to calculate the curve of change of signal power at the input of the receiver as the satellite moves through its orbit, using formulas (7.15) and (7.17).

The results of calculations are presented in Figure 7.12, where the horizontal axis carries quantities representing the position of the satellite in orbit: time read from the perigee and altitude of the satellite over the surface of the Earth (d_{min}). The vertical axis shows the change in signal power at the reception point; we use here the signal power at the point beneath the satellite at the moment when the satellite passes through its apogee as the origin of the reading (0 db point). /173

The two upper curves (solid and dot-dash) show the change in power of the signal at the point beneath the satellite for $\gamma = 0$ and $\gamma = \pm 3°$ respectively.

The set of remaining curves shows the change in power of the satellite at the edges of the zone of coverage for various values of γ.

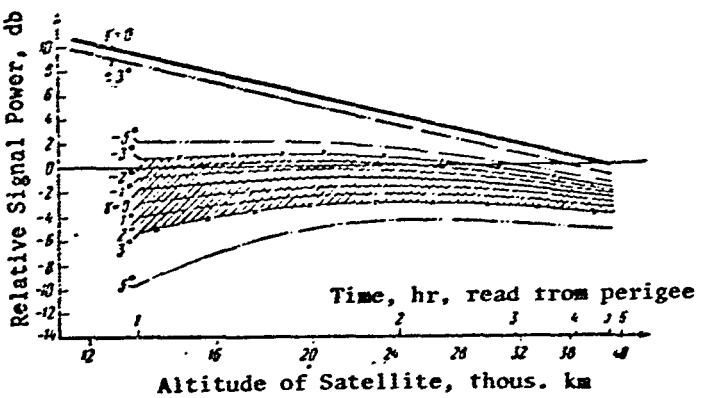


Figure 7.12. Change in Signal Level with Movement of Molniya-1 Satellite

Figure 7.12 indicates that with the value of $\gamma = 3°$ selected, the range of satellite altitudes between 16,000 and 40,000 km (i.e., during a 9.5 hr communications session), the attenuation of the signal is not over 4 db, the quantity characteristic for the initial point in the orbit -- the apogee, where the losses due to enevenness of the radiation pattern amount to $L_\sigma = 3$ db.

Physically, this results from the fact that the attenuation of the signal at the edges of the radiation pattern of the satellite antenna are compensated by the decrease in slant range.

Where $\gamma > 3°$, the compensation conditions are disrupted and the signal losses increase, particularly in the area of low satellite altitudes.

Any attempt to reduce the losses in signal power resulting from unevenness of gain of the satellite antenna requires synthesis of a special radiation pattern.

Generally, in order to assure constant signal power flux density at the surface of the Earth with noncircular orbits, it is necessary to synthesize a radiation pattern not only of special form, but also of variable form, matched to the orbital parameters.

Therefore, in the general case the optimal form of the radiation pattern must be calculated from formula (7.17), on the assumption that $P_{sn} = \text{const}$. /174

However, for some types of noncircular orbits, the task of synthesis of the optimal radiation pattern can be simplified if we do not require constancy of flux within the limits of the entire zone serviced and satisfy ourselves with constancy of power flux only at the edges of the zone serviced, i.e.

$$P_{sn} = \frac{G_{trs}[\phi(t), \gamma(H)]}{d^2_{max}(t)} = \text{const.} \qquad (7.18)$$

This compromise requirement can be met without using variable radiation patterns. Let us show this on the example of the same orbit of the Molniya-1 satellite. The dependences $\phi(t)$ and $d_{max}(t)$ for this orbit are presented in Table 7.1; let us assume for example that the dependence $\gamma(H)$ is as shown in Table 7.2.

TABLE 7.2

H, thous. km	10	15	20	30	40
γ, deg	7	5	3	2	2

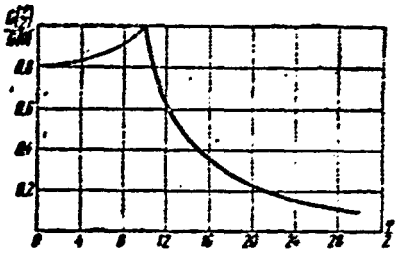


Figure 7.13. Optimal Shape of Radiation Pattern of Satellite Antenna for Satellites in Molniya-type Orbit

Formulas (7.15) and (7.18) can be used to calculate the quasi-optimal shape of the pattern. The result of calculation is shown in Figure 7.13. With this shape of radiation pattern, the attenuation of the signal in the 10-40,000 km altitude range of the satellite is not over 4 db, whereas with the standard radiation pattern and other values the same, the attenuation of the signal γ at low satellite altitudes would reach 15-17 db (see Figure 7.12).

Thus, the use of the quasi-optimal shape of the satellite antenna radiation pattern allows the same losses to be achieved as for circular orbits (L_s = 3 db).

We should note that the losses studied in this chapter cannot be statistically considered, since the statistics of the errors of orientation γ are unknown. Therefore, in performing power calculations for lines, one should consider the maximum values of losses L_s = 3 db, assuming that these losses may occur 100% of the duration of a communications session for certain ground stations. /175

§ 7.7. Determination of Signal Losses in the Atmosphere

When communications are conducted between a space object and the surface of the Earth, the radio waves must pass through space and the atmosphere of the Earth, including the troposphere and the ionosphere. Whereas the space medium, through which the main portion of the path lies, has practically no influence on the propagation conditions of radio waves, the terrestrial atmosphere can radically change these conditions and make them essentially different from the conditions of radio wave propagation in free space.

In the band of frequencies used for satellite communications, the influence of the atmosphere is manifested as the effects of the following factors:

-- attenuation (absorption) of radio waves in the troposphere and ionosphere;

-- reflection and scattering due to heterogeneities encountered along the propagation path;

-- curvature of the trajectory of the radio beam as a result of refraction;

-- change in shape and plane of polarization of radio waves (Faraday effect);

-- appearance of interference resulting from thermal radiation of the atmosphere and absorption noises.

Each of these factors leads in the final analysis to a decrease in the energy (or level) of the signal at the point of reception and therefore should be considered with the required accuracy. In this section, we will study the first factor -- attenuation of the signal due to absorption in the atmosphere.

Radio waves experience the greatest absorption in atmospheric gasses, primarily in oxygen and water vapor[1]. This absorption is selective (frequency dependent), the greatest absorption being observed in frequencies corresponding to the frequencies of intra-molecular transitions.

Thus, the electrons of oxygen molecules, which have constant magnetic moment, may be converted from lower energy levels to higher energy levels by the magnetic field of a wave [7.2]. When this happens, energy is absorbed from the radio wave and the field intensity is correspondingly reduced. Oxygen has two absorption bands, the centers of which correspond to frequencies of 60 GHz and 120 GHz.

Water molecules have constant electrical moment; under the influence of the electric field of a wave, oscillations are excited in them, and are maintained by the energy of the signal. /176
The greatest amplitude of oscillations and, consequently, the greatest absorption of energy, is observed at frequencies of 24 GHz and 165 GHz.

These resonant absorption frequencies are quite high; therefore, in the band of frequencies set aside for satellite communications, the frequency properties of the atmosphere change relatively smoothly. The signal loss in the atmosphere can be characterized by absorption factor L'_a, db/km, which depends on the distribution of pressure, temperature and relative humidity along the signal path.

An approximate model of the atmosphere considering this distribution is shown in Figure 7.14 [7.3]; the solid curves relate to the dry summer atmosphere, the dotted curves -- to very moist atmosphere.

The corresponding analytic dependences on altitude h (km) are as follows [7.3]:

---

[1]The other gases, -- nitrogen, carbon dioxide, methane, ozone, etc. -- absorb the energy of a signal approximately 10^3 times less than water vapor.

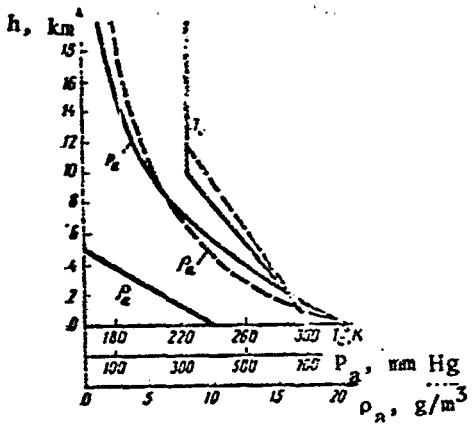


Figure 7.14. Model of the Atmosphere

-- for the temperature of the atmosphere

$T_a = 299°K - 6.5\ h$

where

$0 \leq h \leq 11.6$ km,

$T_a = 223°K$

where

$h > 11.6$ km;

-- for atmospheric pressure

$p_a = 765 \exp(-0.125\ h)$, mm Hg;

-- for absolute humidity

$\rho_a = \rho_0 - 2h(\rho_0 = 10\ g/m^3)$ during periods of low humidity,

$\rho_a = 20 \exp(-0.12\ h)^2/m^3$ where $h \geq 10$ km for periods of high relative humidity.

Since the pressure, temperature and humidity change significantly with altitude, the concept of the characteristic height of the atmosphere [7.4] is introduced for convenience in estimating atmospheric effects. The characteristic height means the height of an atmosphere consisting only of molecules of oxygen or only of molecules of water vapor, distributed evenly with height and introducing the same absorption as the real atmosphere.

As was shown in [7.4], the characteristic height of the atmosphere consisting solely of molecules of oxygen $h_x^1 = 5.3$ km, while the characteristic height of an atmosphere consisting of water molecules $h_x^2 = 2.1$ km.

The path length of a radio signal in the atmosphere obviously depends not only on the thickness of the atmosphere but also on the elevation of the ground station antenna. For $\beta = 90°$, the beam intersects the atmosphere through the shortest distance; as $\beta \to 0$, the length of the signal path in the atmosphere increases significantly (by up to 80 times). The height of the ground station over sea level h_g also has some influence. /177

The equivalent path length of a signal in the atmosphere can be determined from the formula

$$l_a = (h_x - h_g)\operatorname{cosec}\beta. \qquad (7.19)$$

In the interpretation used above, the total absorption of radio waves in the atmosphere can be determined from the formula

$$L_a = L'_{O_2} l_1 + L'_{H_2O} l_2 + L'_{\text{ж}} l_3, \qquad (7.20)$$

where L'_{O_2} and L'_{H_2O} are the absorption factors for oxygen and water vapor at the surface of the Earth.

According to (7.19)

$$l_1 = (h_x^1 - h_3)\operatorname{cosec}\beta,$$
$$l_2 = (h_x^2 - h_3)\operatorname{cosec}\beta,$$
$$l_3 = (h'_{\text{ж}} - h_3)\operatorname{cosec}\beta.$$

The numerical values of the absorption factors L'_{O_2} and L'_{H_2O} are presented in Figure 7.15, which indicate that in the wave range used for satellite communications, the primary losses (in a quiet atmosphere) result from absorption of radio waves by oxygen. /178

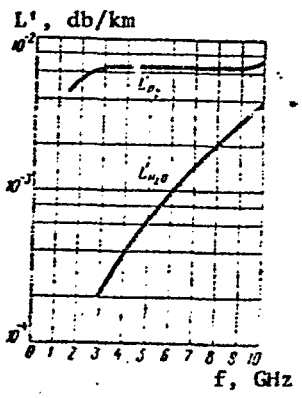


Figure 7.15. Absorption Factors at Sea Level for Oxygen and Water Vapor

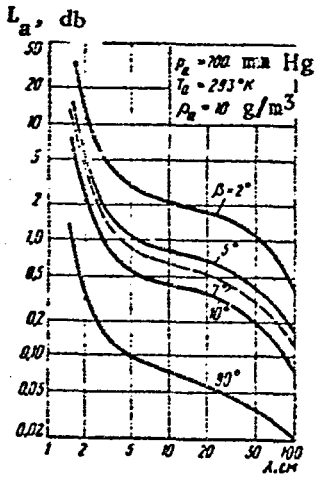


Figure 7.16. Losses in the Atmosphere for Various Elevations

Figure 7.16 shows the frequency dependence of signal losses in a quiet, dry atmosphere on elevation of the antenna, calculated using the formula presented above, namely (7.20), under the condition that $L'_r = 0$. Figure 7.16 indicates that in the 4-6 GHz band, the maximum losses will not exceed 1 db with elevations $\beta \geq 5\text{-}7°$.

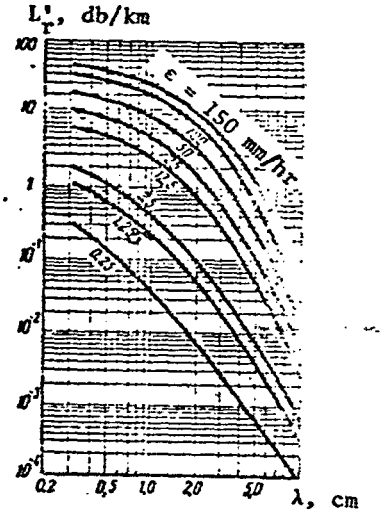


Figure 7.17. Absorption Factor for Rain of Various Intensities

Significant signal attenuation may be caused by wet weather (rain, hail, snow, fog), primarily due to absorption and scattering of electromagnetic energy by water droplets. With a clearly outlined band of precipitation, a portion of the energy may also be reflected back toward the transmitter.

The greatest attenuation is introduced by liquid precipitation (rain, fog); the attenuation resulting from solid precipitation (hail and snow) is significantly less; finally, the presence of suspended particles -- aerosols -- has practically no influence on absorption of the signal and can be ignored under ordinary conditions.

The attenuation of the signal in rain depends on the intensity of the rain and the thickness of the rain zone. Figure 7.17 shows the frequency characteristics of absorption factor L'_r in rain of various intensities ε, from which we can see that with heavy rain (ε = 100-150 mm/hr), the absorption factor at frequencies of 4-6 GHz reaches 0.5-1 db/km. However, rain of this intensity is quite rare, and the length of the area of precipitation of rain of this intensity is generally slight, not over 5-10 km, with a vertical thickness of up to 2 km. Areas of drizzling rain, with intensities of 3-5 mm/hr, on the other hand, may reach 100-300 km in width, with thicknesses of 1-2 km. The length of a signal path within this type of rain zone is calculated using formula (7.19) for h_r = 2 km and $\beta = 5°$, and may reach 25 km.

Figure 7.18 shows the integral curves of the distribution of intensity of precipitation, constructed using the averaged results of observations performed at several points over the USSR. Curve 1 shows the mean annual intensity of precipitation in Moscow, curves 2, 3 and 4 -- in Novosibirsk, Khabarovsk and Tashkent. Curves 5

and 6 define the maximum and minimum mean monthly intensity of precipitation respectively. It follows from this figure that /179
with a probability of 99.9%, the greatest mean monthly intensity of precipitation is not over 3 mm/hr and, consequently, the absorption, calculated using the curves of Figure 7.17 and formula (7.19), (7.20) for $\beta \geq 5°$, will not exceed 0.5 db.

The next most significant absorber of radio wave energy is fog. The intensity of fog is measured by the maximum optical visibility S in meters, its absorbing capacity L_F' (in db/km) is determined by the absolute moisture content ρ, g/m^3.

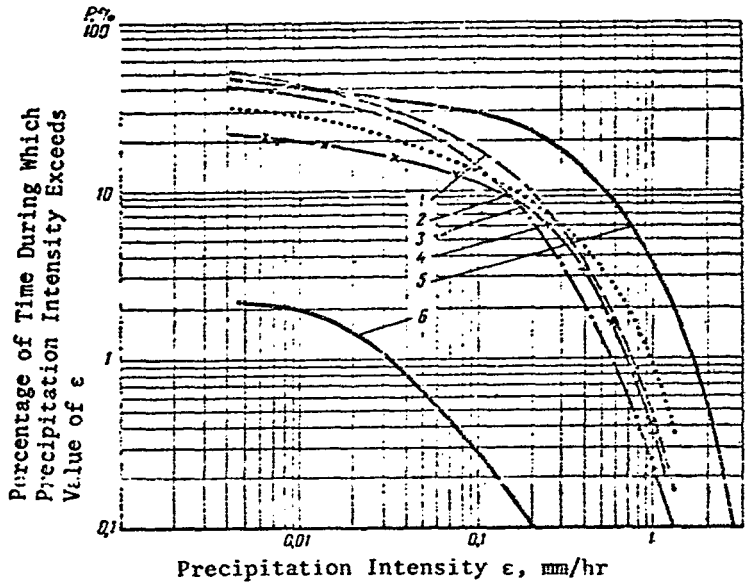


Figure 7.18. Integral Distribution of Precipitation Intensity for Certain Points in the Territory of the USSR

The relationship of these parameters can be expressed by the empirical formulas [7.5]:

$$\rho = 3S^{-1,3}, \quad L_F' = 0{,}483\frac{\rho}{\lambda^2}.$$

The mean vertical extent of a fog area is slight, 0.5-0.6 km, while the horizontal extent may reach 100 km, and the duration of existence of fog may be significantly greater than that of rain.

Thus, although the absolute value of absorption in fog L_F', db, is relatively slight, the great times of existence of this absorption force us to consider it in calculating lines. /180

Figure 7.19 shows curves [7.6], allowing us to determine the statistical value of absorption in fog and clouds at 4 GHz for the climate of southern England. They indicate that during 99.9% of the time, the attenuation of the signal with $\beta = 5°$ will not exceed 1 db. Considering the climatic peculiarities of southern England (abundant fog and dense clouds), we can assume that this absorption will be observed significantly less frequently in the USSR and that the general statistical characteristics of absorption in fog will be approximately the same as in rain.

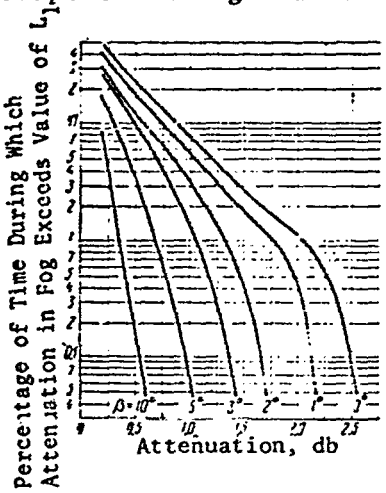


Figure 7.19. Integral Distribution of Losses in Fog for Various Angles of Elevation

The ionosphere also influences the conditions of transmission of radio waves. However, absorption in the ionosphere at frequencies of over 1,000 MHz is quite slight and, according to formula (7.6),

$$L_i \approx \frac{2500}{f^2_{MHz}}, \text{ db},$$

does not exceed $2.5 \cdot 10^{-3}$ db, even at low propagation angles.

§ 7.8. Losses Due to Refraction

Another type of loss of signal energy is related to the presence of the atmosphere: losses due to refraction. If the terrestrial atmosphere had identical parameters at all points, including identical dielectric constant, i.e., in other words, if it were homogeneous, a radio beam passing through the atmosphere, like a light beam, would follow a straight line path. However, since the atmosphere is actually quite heterogeneous and its dielectric constant changes as a function of altitude over the surface, humidity and temperature, the trajectory of a radio beam passing through areas with different values of dielectric constant will bend. This phenomenon is called refraction. The refraction of radio waves is observed both as they pass through the ionosphere, and as they pass through the troposphere; however, the contribution of these areas of the atmosphere to the total refraction in the frequency range in question is quite different.

Thus, the change in dielectric constant in the ionosphere results completely from the influence of electrons and ions. The influence of neutral particles on dielectric constant is quite negligible and is considered only in calculating the number of collisions between electrons and ions. /181

As the results of calculations and experiments have shown, the value of ionospheric refraction is inversely proportional to the square of frequency [7.7] and independent of changes in electron density with time and location, which reach their maximum value in tropical regions. The contribution of ionospheric refraction to total refraction in the band of frequencies on 0.1 GHz and higher is not great, and can be calculated from the following formula [7.8]:

$$\delta_i = \frac{57.3 \cos \beta}{f^2 \sin^3 \beta}, \text{ degrees.}$$

In the lower layers of the atmosphere, in the troposphere, the significance of electrons and ions is slight and the value of dielectric constant is completely determined by neutral particles (molecules). Tropospheric refraction in the frequency range in question is independent of frequency. For ground antenna elevation angles β differing slightly from 0, the value of tropospheric refraction can be calculated by the formula [7.9]

$$\delta_{tr} = (n - 1) \cot \beta, \text{ degrees,}$$

where $n = \sqrt{\varepsilon_a}$ is the index of refraction of the troposphere;

ε_a is the dielectric constant of the atmosphere.

In the overwhelming majority of cases, the change in n with altitude h over the surface of the Earth follows the exponential rule

$$n(h) = 1 + a \exp(-bh),$$

where a and b are constants which can be determined statistically for various climatic conditions.

At the present time, the ICCR uses the following altitude dependence of the refractive index as the basic initial model of the atmosphere [7.10]:

$$n(h) = 1 + 289 \cdot 10^{-6} \exp(-0.136h).$$

The value of refraction calculated for this model of the atmosphere is generally called regular refraction. However, the exponential dependence n(h) is not always observed; under certain weather conditions in certain areas of altitude, significant deviations may occur. However, they are observed comparatively rarely; most frequently, actually practically always, comparatively slight fluctuations in n occur about this exponential dependence.

The refraction resulting from the presence of fluctuations in the index of refraction of the medium is generally called irregular refraction. /182

The reason for these fluctuations in n are turbulent movement of the air, resulting from unevenness of the surface of the Earth and unevenness of its heating, and comparatively stable stratified formations, resulting from the influence of terrestrial gravity and variations in air density at various altitudes over the surface of the Earth.

The value of the index of refraction of the air changes not only in the vertical direction, but also in the horizontal direction along the surface of the Earth. The change in n in the horizontal direction results from a sharp change in weather conditions or, as meteorologists say, from the passage of warm or cold fronts. The greatest horizontal changes in n are observed at the ocean shoreline.

Thus, weather conditions and, consequently, the index of refraction of the air, are not constant with time; therefore, the refraction conditions also change with time. However, due to the relative slightness in fluctuations of index of refraction of the medium, the mean square value of the angle of irregular refraction is generally an order of magnitude less than the mean square value of the angle of regular refraction and therefore the error in angle due to irregular refraction is ignored.

Figure 7.20 shows the dependence of angles of tropospheric and ionospheric refraction on frequency for various elevation angles, calculated using the formulas presented above [7.8]. As follows from the figure, when frequencies above 1000 MHz are used, summary refraction $\delta_r = \delta_p + \delta_{tr}$ is determined primarily by the tropospheric refraction and is independent of frequency.

Although the error in refraction for elevation angles exceeding 5° is not great (on the order of 10 minutes), in satellite communications systems it may be comparable to the width of the radiation pattern of ground station antennas. Therefore, in

guiding and tracking a ground station antenna using a program which ignores refraction, the satellite may be "lost," if the width of the radiation pattern of the ground station antennas is less than the angle of refraction.

As we know, most existing ground stations use the method of automatic guidance of ground station antennas based on the signal of the satellite plus the method of programmed tracking, which is more economical both from the standpoint of cost and from the standpoint of equivalent noise temperature of the system. /183

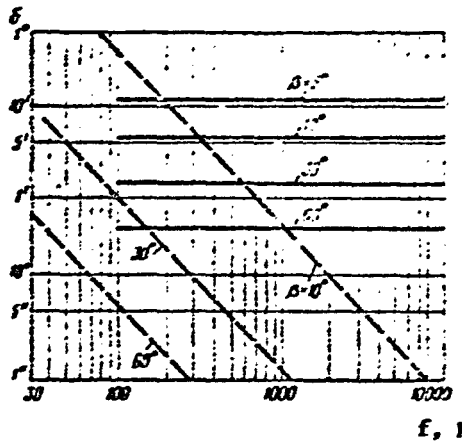


Figure 7.20. Dependence of Tropospheric (solid lines) and Ionospheric (dotted lines) Refraction on Frequency at Various Elevation Angles

However, programmed tracking requires precise calculation of the ephemerides of the satellite, which are used as input information for the digital control device, and the errors related to refraction of the signal in the atmosphere, which are functions of elevation.

According to [7.11], the errors due to refraction with elevations on the order of 5° at frequencies of 4-6 GHz can be decreased to 20-40 seconds using the following two methods of correction:

1) measurement of refraction for a fixed model of the troposphere and calculation of the correction angle as a function of elevation angle;

2) prediction of refraction as a function of location, time of day and time of year for a fixed troposphere model and calculation of the correction angle as a function of elevation angle.

Thus, refraction errors can be calculated, recorded in the digital programming device and corrections made as functions of the elevation of the antenna in real time. Knowing the width of the radiation pattern of the ground station antenna and using Figure 7.8, we can calculate the losses of signal due to refraction errors. For example, for a parabolic antenna with a diameter of 30 m with a radiation pattern width at the half power level of 0.2° at 4 GHz, the losses of signal power due to refraction at elevations $\beta = 5\text{-}7°$ can reach 10 db. Correction of refraction errors by one of the methods outlined above can reduce these losses to 0.05-0.1 db.

§ 7.9. The Faraday Effect and Losses Due to Antenna Polarization Mismatch

In studying the process of propagation of radio waves through the ionosphere, we must consider the fact that the radio waves are traveling through an ionized gas in the permanent magnetic field of the Earth (at the equator it is approximately equal to 20 A/m, at the poles it is double this value), the presence of which gives the ionized gas the properties of an anisotropic medium, i.e., a medium, the properties of which depend on the direction of propagation.

In the general case, i.e., with arbitrary orientation of the direction of propagation relative to the direction of the Earth's permanent magnetic field, a linearly polarized wave propagating through such a medium is split into two elliptically polarized components (normal wave): the ordinary and extraordinary components. The conditions of propagation of these components and, particularly, their phase velocities in the ionosphere, are different; therefore, as they travel a certain distance a phase shift arises between them, resulting in rotation of the plane of polarization of the resultant wave, referred to as the Faraday effect. In the literature, this effect is frequently called the "rotary" Doppler effect. /184

In order to explain the physics of this phenomenon, let us analyze the propagation of a wave in the permanent magnetic field of the Earth, which has only a longitudinal component, since it is this component which causes the rotation of the plane of polarization of the wave.

As we know, as they propagate in a longitudinal magnetic field, electrons receive a rotary movement in the plane perpendicular to the direction of propagation of the wave. Therefore, when a linearly polarized wave is broken down into two components, the direction of rotation of the resulting field vector corresponds to the direction of rotation of the electrons for only one component; for the other component, the directions are directly opposite.

Naturally, the conditions of propagation of these components differ, leading to the effect described. When communications are conducted through a relay station on an Earth satellite, and when antennas with linear polarization are used for transmission and reception, the presence of this effect can result in significant polarization fading of the signal due to the continuous rotation of the plane of polarization of the wave, since when the satellite moves, both the angle between the direction of propagation and the magnetic field of the Earth and the path length of propagation of the radio wave in the ionosphere change.

According to [7.12], the angle of rotation of the plane of polarization of the wave for frequencies over 200 MHz for the case of quasilongitudinal propagation, i.e., where

$$\frac{4\pi e^2}{m\omega^2}N=\frac{\omega_0^2}{\omega^2}\ll 1 \quad \text{and} \quad \frac{\omega_H^2}{\omega^2}\ll 1,$$

is equal to

$$\Psi_\phi=\frac{\omega_0^2\omega_H^2}{2c\omega^2}\left\{(\cos\gamma_0+\sin\gamma_0\operatorname{ctg}\beta)\int_0^H NdH-\right.$$
$$\left.-\operatorname{ctg}\beta\frac{\sin\gamma_0}{\sin^2\beta}\left[\int_0^H\frac{NH\,dH}{R_e}-\frac{3-\sin^2\beta}{2\sin^2\beta}\int_0^H\frac{NH^2\,dH}{R_e}\right]\right\}, \qquad (7.21)$$

where ω_0 is the angular plasma frequency;

ω_H is the angular gyroscopic frequency of an electron

ω is the angular working frequency;

c is the velocity of light;

m and e are the mass and charge of an electron;

γ_0 is the angle between the direction of the wave and the magnetic field of the Earth; /185

H is the altitude of the satellite;

N is the electron concentration of the ionosphere;

R_e is the radius of the Earth.

Thus, if we know the rule of change of electron concentration with altitude N(h), we can calculate the angle of rotation of the plane of polarization.

Figures 7.21-7.23 show the dependences of angle of rotation of the plane of polarization on working frequency, altitude and satellite zenith angle taken from [7.12].

Figure 7.21 shows Ψ_ϕ as a function of frequency for $\gamma_0 = 30°$ and H = 1500 km.

The rule of change of electron concentration for this case according to [7.13] is

$$N = N_{max} \exp\left[-k_i(h - h_1 - h_m)\right], \qquad (7.22)$$

where N_{max} is the maximum electron concentration in the F_2 layer;

h_1 is the height of the lower boundary of the ionosphere;

h_m is the thickness of the lower half of the ionosphere (from its lower boundary to the height of the maximum electron concentration in the F_2 layer);

k_i is a factor which is defined on the basis of ionization density at altitudes of over 300 km ($k_i = 3.8 \cdot 10^{-3}$ km^{-1}).

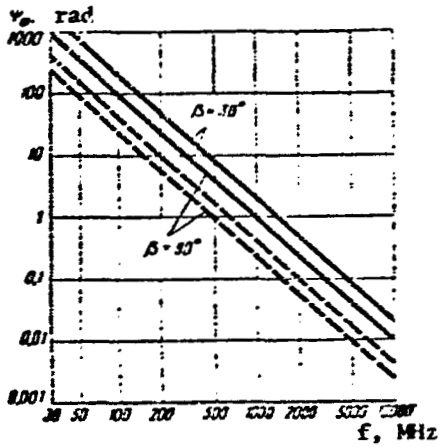


Figure 7.21. Frequency Dependence of Angle of Rotation of Plane of Polarization of Signal

The solid lines on the figure show the curves for f_0 = 15 MHz, the dotted lines -- for f_0 = 7 MHz.

Figure 7.22 shows the change in angle ψ_ϕ with altitude of the spacecraft; the solid lines relate to the case ϕ_0 = 15 MHz, the dotted lines -- ϕ_0 = 7 MHz. The curves are calculated for γ_0 = 30° and f = 1000 MHz for several values of satellite elevation angles. As we can see from the graph, where H > 500 km the angle of rotation of the plane of polarization changes little with altitude.

Figure 7.23 shows ψ_ϕ as a function of elevation of the satellite with values of angle γ_0 = 0°, 30°, 90°, corresponding to curves 1, 2 and 3 for H = 5000 km and f = 500 MHz.

Thus, the presence of the Earth's permanent magnetic field leads to the fact that as a linearly polarized wave propagates through the ionosphere, the resulting field is generally elliptically polarized, with the orientation of the large half axis of the ellipse of polarization dependent on the direction of propagation, frequency and path length. /186

Consideration of the angle (γ_0) between the direction of propagation of a wave and the magnetic field of the Earth is very important in calculating the angle of rotation of the plane of polarization of the summary wave. Depending on the quantities included in formula (7.21) the angle of rotation of the plane of

polarization may vary over rather broad limits, which, in relationship to the movement of space objects using linearly polarized antennas, leads to deep polarization fading of the signal received, following the rule $P_s(t) \approx P_0 \cos 2\psi(t)$.

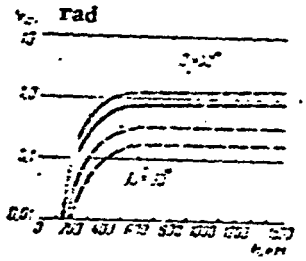


Figure 7.22. Angle of Rotation of Plane of Polarization as a Function of Satellite Height

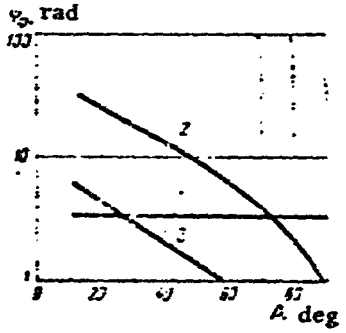


Figure 7.23. Rotation of Plane of Polarization as a Function of Antenna Elevation Angle

The magnitude and rate of fading are inversely proportional to the square of frequency.

Fading may be so deep that at certain moments, a significant reduction in the level of the signal at the receiving end is experienced. Therefore, in calculating the signal power loads of a satellite communications link, this phenomenon must be considered and the required power reserve for polarization losses introduced. The value of this reserve depends on the type of polarization used for the satellite and ground station antennas.

According to [7.8], the expression for normalized power transmission factor between two antennas in the most general case (both antennas with elliptical polarization) can be written in the following form

$$K=\frac{1}{2}\left[1+\frac{4e_1e_2}{(1+e_1^2)(1+e_2^2)}+\frac{(1-e_1^2)(1-e_2^2)}{(1+e_1^2)(1+e_2^2)}\cos 2\psi\right], \qquad (7.23)$$

where e_1 and e_2 are the ellipticity factors (ratio of semiminor to semimajor axis of ellipse) of polarization of the transmitting and receiving antennas respectively; /187

ψ is the angle between the semimajor axes of the polarization ellipses of the transmitting and receiving antennas.

Knowing the relative positions and parameters of the ellipses of polarization of the transmitting and receiving antennas, we can use formula (7.23) to determine the signal power losses L_p for a given receiving antenna in comparison to an antenna, whose ellipse of polarization coincides completely with the ellipse of polarization of the transmitting antenna.

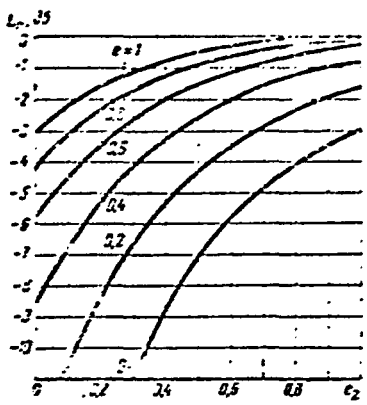


Figure 7.24. Losses Due to Ellipticity of Polarization of Receiving and Transmitting Antennas

Figure 7.24 shows the curves of maximum losses for $\psi = 90°$ for various values of ellipticity of polarization, while Table 7.3 shows the values of transmission factor for various combinations of polarizations of transmitting and receiving antennas under the condition that each of the antennas has either linear or circular polarization.

As follows from Table 7.3, with linear polarization of the transmitting and receiving antennas, the transmission factor is at its maximum when the two polarization vectors are collinear.

Obviously, when communications are conducted through a moving satellite, the conditions of linearity of vectors will be disrupted due to the change in mutual position of the objects in communication, and also due to /188 the influence of the medium on the radio waves propagating through it. Therefore, the necessity arises of continually adjusting the position of the vector of the receiving antennas as the vector of the transmitted wave deviates, which is quite difficult.

It is preferable to give both the satellite and ground station antennas circular field polarization with identical direction of rotation of the polarization vector.

However, even with this version, design difficulties arise, consisting in that the field of circular polarization cannot be produced over the entire width of the radiation pattern, but only in its central portion. It is pointed out in [7.14] that the maximum width of the radiation pattern of modern antennas in which polarization near circular can be achieved is not over 100° (the angle of visibility of the Earth from an altitude of approximately 2000 km). The use of circularly polarized antennas for altitudes of less than 2000 km leads to additional signal losses at the edges

of the radiation pattern due to polarization mismatch; the less this altitude, the greater the signal loss. For altitudes such that the losses exceed 3 db, there is good reason to use an antenna with linear field polarization on the satellite, since losses due to polarization mismatch between the satellite and ground antenna in this case will not exceed 3 db and do not change as the satellite moves.

TABLE 7.3

Polarization of Receiving Antenna	Transmission Factor for Various Polarizations of Transmitting Antenna		
	Linear	Circular with Right Rotation of Polarization Vector	Circular with Left Rotation of Polarization Vector
Linear	$\frac{1}{2}(1 + \cos 2\psi)$	1/2	1/2
Circular with Right Rotation of Polarization Vector	1/2	1	0
Circular with Left Rotation of Polarization Vector	1/2	0	1

It should be added that the only factor preventing use of antennas with circular polarization at the transmitting and receiving ends for high altitude satellites is the effect of changing direction of rotation of a circularly polarized wave as it passes through the ionosphere. However, as results of calculations performed in [7.8] have shown, when frequencies over 100 MHz are used, this effect practically never appears.

Thus, for communications lines using high altitude satellites, it is desirable to use circularly polarized antennas. The signal power losses resulting from the ellipticity of polarization arising in practice can be easily calculated using formula (7.23) or Figure 7.24, which indicate that for $e_1 = e_2 = 0.8$, the value of L_p will not exceed 0.25 db.

Statistical estimation of polarization losses is impossible; therefore, in calculating the transmission levels for a communications line they should be considered constant, i. e., existing 100% of the time.

§ 7.10. Phase Dispersion of Signals in the Atmosphere

The influence of the atmosphere is also manifested in the creation of relative phase shift between various frequency components of a signal as it passes through near-Earth space, particularly through the ionosphere. Like Faraday rotation, the magnitude of these shifts is inversely proportional to the square of frequency; they result from the frequency dependence of the refractive index of the atmosphere. /189

As was shown in [7.15], the refractive index of the atmosphere in the decimeter and centimeter wave bands is determined by the formula

$$n \approx 1 - \frac{f_0^2}{2f^2},$$

where f_0 is the plasma frequency.

The value of the plasma frequency depends on the concentration of electrons in the medium and, consequently, is subject to daily and seasonal changes, as well as the influence of solar activity. For a regular ionosphere, f_0 may vary from 3 to 15 MHz, the most characteristic value of f_0 being 7 MHz.

The total phase shift of the signal as it passes through the transmitter and receiver

$$\varphi_0 = \frac{\omega}{c} \int n \, dl,$$

where $\omega = 2\pi f$ and f is the signal frequency;
c is the speed of light;
l is the path length of the signal.

From this, the group delay time of the signal

$$\tau = \frac{d\varphi_0}{d\omega} = \frac{d}{c} + \frac{1}{2cf^2} \int f_0^2 \, dl,$$

where d is the distance between the transmitter and receiver.

The approximate difference in the group delay time $\Delta\tau$ of the extreme components of the wide band signal with band Δf will be equal to:

$$\Delta\tau = \frac{\Delta f}{cf^2}\int f_0^2\, dl.$$

In order to avoid distortions in the wide band signal as it passes through the ionosphere, the following conditions should be met:

$$\Delta\tau\Delta f = \frac{\Delta f^2}{cf^2}\int f_0\, dl \ll 1.$$

The integrand characterizes the state of the ionosphere; therefore, its value generally depends on changes in f_0. In order to produce a quantitative estimate of the bandwidth of the ionosphere, we assume in the first approximation

$$\int f_0^2\, dl \approx \tilde{f}_0^2 l_i,$$

where $\tilde{f}_0$ is the mean value of the plasma frequency of the ionosphere;

l_i is the pathlength of the signal in the ionosphere.

Then

$$\Delta\tau\Delta f \approx \frac{\Delta f^2}{cf^2}\tilde{f}_0^2\, l_i.$$

Fixing the values $\tilde{f}_0$ = 7 MHz, $l_{i\ max}$ = 1000 km and assuming /190
$\Delta\tau\Delta f$ = 0.1, it is not difficult to determine that the greatest radio signal bandwidth which can be transmitted without distortion is

$$\Delta f < \sqrt[3]{3\cdot 10^{-11} f^3}$$

The results of the calculations are presented in Table 7.4, from which we see that the bandwidth of the atmosphere is not over

25 MHz at 1000 MHz, reaching 500 MHz in the 4-6 GHz band. With strong perturbations of the ionosphere, particularly during years of the maximum solar activity, the value of $\Delta\tau$ may increase by two to three times, leading to a proportional reduction in the band of frequencies of the ionosphere.

TABLE 7.4

f, MHz	500	1000	5000	10000
Δf, MHz	10	25	270	750

These limitations should be kept in mind in planning multichannel television and telephone communications systems.

We should note that in this calculation we have considered only the influence of the ionosphere, since the value of f_0^2 for the other layers of the atmosphere is approximately 100 times less than for the ionosphere, so that the bandwidth of the remaining portion of the atmosphere is almost 5 times greater that that of the ionosphere.

§ 7.11. Receiver System Noises

Let us now go over to the last stage in our calculations -- determination of the total noise power created at the input of the receiving system by various sources. Precise and exhaustive evaluation of this quantity is of great significance in principle, since the noise level requirement for the receiver system in the final analysis is determined by the minimum value of the signal which can be received with the required quality.

As was indicated in § 7.2, the noise power at the input of the receiver can be calculated using the formula

$$P_n = kT_\Sigma \Delta f_n \qquad (7.24)$$

Let us now study the quantities included in this formula. The equivalent noise band Δf_n (power band) of the receiver, in turn, is determined by the expression

$$\Delta f_n = \frac{1}{K^2(f_0)} \int_{-\infty}^{\infty} K^2(f)\,df$$

(where K(f) is the frequency characteristic of the receiver; f_0 is the center frequency of the receiver) and essentially /191 represents the bandwidth of the equivalent receiver with rectangular frequency characteristic, creating the same noise power at the output as the actual receiver with frequency characteristic K(f). The equivalent noise band of the receiver is related to the band of frequencies bounded by the -3 db level, according to the formula

$$\Delta f_n = \gamma \Delta f_{0.7},$$

in which the value of coefficient γ characterizes the selectivity of the receiver. For particular cases, when a receiver contains n identically adjusted stages, these values (γ_1 for single-circuit and γ_2 for dual-circuit stages) are presented in Table 7.5 [7.16].

Usually in planning communications lines, either the equivalent noise band of the receiver or the resulting band of frequencies, determined by the form and parameters of the signal received, is fixed; in the latter case, the transition from $\Delta f_{0.7}$ to Δf_n can be performed using Table 7.5.

TABLE 7.5

n	1	2	3	4	5	10
γ_1	1.57	1.22	1.13	1.13	1.11	1.09
γ_2	1.11	1.038	1.022	1.014	1.009	1.002

The next quantity included in formula (7.24) is the equivalent noise temperature of the receiver system T_Σ. Physically, this quantity shows the temperature to which we must heat or cool a certain active resistance, equal to the input impedance of the receiver in order that the power of fluctuating noise measured across this resistance will be equal to the power of all noise in the system measured in the same frequency band (usually 1 Hz).

The total noise temperature of the entire receiver system, consisting of an antenna with its radiator, the antenna-waveguide system and the receiver itself, is equal to

$$T_{\Sigma} = T_A + T_0 \frac{(1 - \eta_w)}{\eta_w} + \frac{T_r}{\eta_w}, \qquad (7.25)$$

where T_A is the equivalent noise temperature of the antenna;

T_0 is the absolute temperature of the medium, 290 °K;

T_r is the equivalent noise temperature of the receiver itself, resulting from its internal noises;

η_w is the transmission factor of the waveguide system.

This formula defines the resultant noise temperature, calculated for the input of the antenna. It is not difficult to note that this temperature depends on the transmission factor of the waveguide and each decibel of losses in the waveguide is equivalent to an increase in temperature of 73°K. Actually, /192

$$T_w = \frac{293(1-0.8)}{0.8} \approx 73°\text{K}.$$

It is important to emphasize that as losses increase in the waveguide, the signal/noise ratio at the input of the receiver rapidly drop, since in addition to attenuating the energy of the signal, the noise power is increased. Therefore, losses in the waveguide should be reduced to 0.1-0.2 db.

The noise temperature of the receiver T_r depends on the type and design of its input circuit. These problems will be discussed in detail in the second volume. Here, for a general estimation, we present only an approximate comparative characterization of various types of low-noise input circuits. Figure 7.25 shows the values of noise temperature which can be achieved by transistors (1), diode mixers (2), tunnel diodes (3), TWT (4), electron tubes (5), masers, cooled with liquid nitrogen (6), parametric amplifiers (7) and masers cooled with liquid helium (8). The figure indicates that when modern low-noise input devices are used (parametric amplifiers and masers), the noise temperature of the receiver can be reduced to 10-50°K. The limit of increasing sensitivity of receivers is set by the fluctuating noise accompanying the reception of discrete quanta.

As calculations have shown, in the centimeter wave band the quantum noise temperature is low, approximately 0.5°K at 10 GHz. However, it increases with frequency, and in the optical band this type of noise becomes very significant.

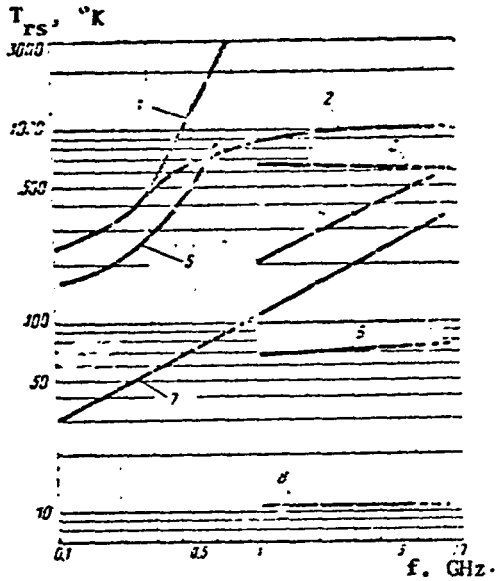


Figure 7.25. Equivalent Noise Temperature of Modern Receiver Devices

Significant difficulties arise in precise determination of the equivalent noise temperature of an antenna, since it depends on a large number of factors. The most important of these are considered in the following expression: /193

$$T_A = T_c + T_a + T_g + T_{ag} + T_{nA} + T_f, \quad (7.26)$$

where T_c is the equivalent antenna noise temperature determined by the reception of cosmic radio radiation;

T_a is the equivalent temperature resulting from the reception of atmospheric radiation;

T_g is the equivalent temperature resulting from the reception of radiation of the surface of the Earth;

T_{ag} is the equivalent temperature resulting from the reception of atmospheric radiation (T_a) reflected from the surface of the ground;

T_{nA} is the equivalent noise temperature of the thermal radiation of the antenna, resulting from optical losses in its elements;

T_f is the equivalent noise temperature resulting from the use of a radome.

Before going over to estimation of the individual components, let us describe our general methodology of calculation of antenna noise temperatures.

According to the general statements of the theory of radiation, if an antenna is in an infinite volume of absorbing medium with homogeneous kinetic temperature T, at thermodynamic equilibrium it absorbs and reradiates power equal to the power of radiation of the medium, and in this case $T_a = T$.

When the temperature of the surrounding medium differs in various directions, which occurs in practice, in order to determine T_a we must multiply these temperatures by the gain of the antenna in the corresponding direction and then average for the entire sphere, i.e.

$$T_A = \frac{1}{4\pi} \int_{4\pi} T_b(\beta, \psi) G(\beta, \psi) d\Omega. \tag{7.27}$$

where $T_b(\beta, \psi)$ is the brightness temperature of radiation in direction β, ψ in the spherical system of coordinates;
$G(\beta, \psi)$ is the gain of the antenna (relative to an isotropic radiator) in the same direction.

The concept of "brightness temperature" is introduced to describe sources of radiation; it is defined as the temperature of an absolutely black body with the same brightness at the given frequency and in the given direction as the source in question.

In order to characterize sources of radiation with uneven distribution of brightness temperature, we utilize the concept of the mean brightness temperature or effective radiation temperature /194

$$T_m = \frac{1}{\Omega_i} \int_{\Omega_i} T_b(\beta, \psi) d\Omega,$$

where Ω_i is the solid angle of the source of radiation.

Sometimes in order to evaluate noise radio radiation, we use the concept of the spectral density of the radiation flux. The spectral flux density is expressed in watts/m^2·Hz and is the flux of power radiated per unit frequency band and passing through a unit area.

The spectral flux density from a source with a small angular dimension is related to the effective temperature by the following equation

$$W = \frac{2kT_m}{\lambda^2} \Omega_i.$$

If a source has large angular dimensions (relative to the width of the antenna radiation pattern), then

$$W = \frac{2kT_b}{\lambda^2}\,\Omega_A,$$

where Ω_A is the width of the radiation pattern of the ground station antenna, *sr*.

Thus, in order to calculate the temperature of an antenna, we must know the distribution of the brightness temperature of the medium surrounding the antenna in solid angle 4π *sr*, and the radiation pattern of the antenna in the same angle.

In practice, two cases are quite frequently encountered:

1. When within the limits of the main lobe of the radiation pattern of the antenna, the brightness temperature does not change or changes little, while the radiation received on the side lobes can be ignored.

Then T_b in formula (7.27) can be brought outside the integral sign and the expression for the equivalent noise temperature of the antenna becomes as follows: $T_A = T_b$, since $\int_{4\pi} G(\beta, \psi) = 4\pi$. This case occurs when the distributed radiation of the galaxy is received by directional antennas.

2. When the angular dimensions of the radiation source are small in comparison with the width of the radiation pattern of the antenna ($\Omega_i \ll \Omega_A$). We can consider in this case that within the limits of the solid angle of the source $G(\beta, \psi) = G_{max} = \text{const}$; then, the expression for the equivalent noise temperature of the antenna becomes:

$$T_A = T_m \frac{\Omega_i}{\Omega_A}. \qquad (7.28)$$

The second case occurs when the radiation of discrete sources is received. /195

The resulting noise temperature of the antenna for practical calculations can be estimated as follows: divide the solid angle around the antenna into a number of areas corresponding to the main, side and rear lobes, calculate the noise temperature

individually for reception on the main, side and rear lobes and add the results.

In order to simplify the calculation, the gain within limits of the rear and side lobes is considered constant. In this case, expression (7.27) can be written as follows

$$T_A = \frac{G_{mn}}{4\pi} \int_{\Omega_{mn}} T_b(\beta, \psi) + \frac{1}{4\pi} \sum_{i=1}^{n} \int_{\Omega_{sdi}} G_{sdi} T_b(\beta, \psi) d\Omega.$$

Expanding this expression considering all noise components represented by formula (7.26) and keeping both of the cases studied above for sharply directed ground antennas in mind, we produce the full formula for calculation of the equivalent noise temperature[1]:

$$T_{Ag} = T_{bk}(\beta) + T_{ba}(\beta) + T_{nA} + T_f(\beta) + c(T_{ba} + T_{ba-g}), \tag{7.29}$$

where c is a coefficient considering the integral energy level of the side lobes in the rear half space:

$$c = \frac{1}{2} \frac{\sum_{i=1}^{n} \int_{\Omega_{sdi}} G_{sdi}(\beta, \psi) d\Omega}{\int_{\Omega_{mn}} G_{mn}(\beta, \psi) d\Omega}.$$

A detailed quantitative estimate of coefficient c for various types of antennas will be presented in the second volume of this book. We note here only that the most typical value of this coefficient is 0.1-0.2.

The values for the satellite antenna are

$$T_{As} = T_{bg} + T_{ba} + T_{ba-g} + 2cT_{bk}. \tag{7.30}$$

---

[1]Symbols same as in formula (7.26). The subscript "b" means that it is the brightness temperature of the source with which we are concerned.

The presence of the coefficient 2 before the last component in formula (7.30) results from the fact that cosmic radiation acts not only on the lobes in the rear half space but on the side lobes of the antenna as well.

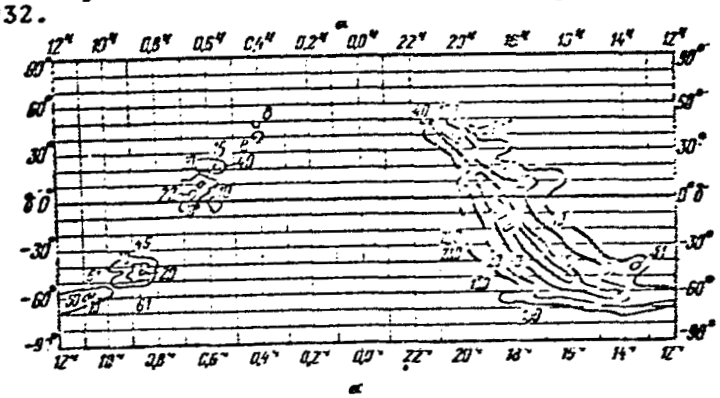
Let us now go over to evaluation of the components of the noise temperature of the antenna presented in formulas (7.29) and (7.30).

§ 7.12. Determination of Brightness Temperature of Radiation Sources

/196

The equivalent noise temperature of cosmic radiation is determined by the radiation of the galaxy, the radiation of the sun, the moon, the planets and other discrete sources, the radiation of the polar auroras, etc. The radiation arriving at the Earth from space and recorded as the background was first detected in 1932.

a)



b)

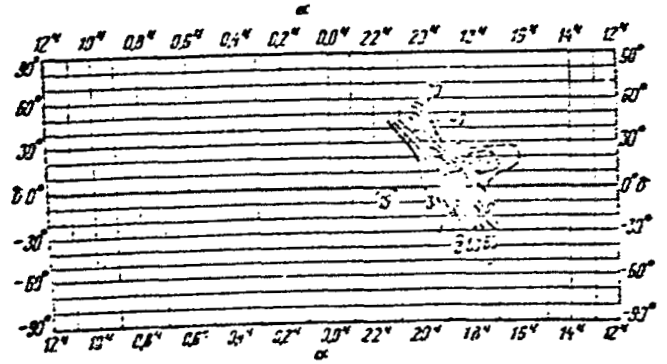


Figure 7.26. Isophotes of Galactic Radiation: a) at 600 MHz, b) at 910 MHz. Key: ч = hr.

Detailed study of the distribution of cosmic radiation and the composition of special isophote maps, showing curves connecting various areas of the sky with identical radiation intensity (Figure 7.26)[1] have allowed us to accumulate a large quantity of data on the radiation of the galaxy. It has been established that this radiation is most intensive in the plane of the galactic equator, and decreases smoothly in the directions of the poles. /197

The maximum of radiation corresponds to the center of the galaxy, in the direction of the constellation Sagitarius. In addition to this primary maximum of radiation, there are a number of secondary maxima, located on the galactic equator. The radiation of the galaxy has a continuous spectrum and is weakly polarized [7.17]. Therefore, when it is received on a polarized antenna, we can consider with good accuracy that the radiation is received at half intensity.

The intensity of the radiation of the galaxy decreases with increasing frequency. Figure 7.27 shows the maximum and minimum values of intensity of galactic radiation, expressed in units of brightness temperature T_{bk} as a function of frequency.

In addition to the reduction in intensity of the radiation of the galaxy with increasing frequency, the dimensions of the area of maximum radiation also decrease. For example, according to [7.8], whereas at 100 MHz the extent of the area, at the boundaries of which the intensity of the radiation drops to half in comparison to the intensity of the radiation at the center of the galaxy, is 80° in the direction along the equator and about 20° in the perpendicular direction, at about 1 GHz the angular dimensions of this area are about 10° x 4°.

A temperature dependence is also observed as concerns the temperature drop of radiation between the center and poles of the galaxy. The higher the frequency, the greater this drop.

The maximum value of equivalent temperature of cosmic noise from the center of the galaxy, as follows from Figure 7.27, is not over 30°K at 1 GHz, and a fraction of a kelvin at 10 GHz.

Thus, when operating at frequencies over 1 GHz, the contribution of galactic noise to the total antenna temperature is insignificant. However, it must be considered, estimating the noise level at the input of the ground station receiver on the basis of the maximum value of the brightness temperature of the galactic noise, since the fixed signal/noise ratio in the communications channel should be retained when the direction of the antenna

---

[1]The numbers on isophotes indicate the absolute temperature in kelvins (α -- right ascension, δ -- declination).

toward the object coincides with the direction toward the area of the galaxy with maximum radiation.

One of the brightest sources of radiation is the sun. The radio radiation of the sun can be divided into the following types: thermal (primary component), slowly changing component (sun spots), /198 storm noise, flares and individual bursts. At frequencies over 20 MHz, the solar radiation consists primarily of thermal and slowly changing noise.

The thermal component corresponds to the radiation of the quiet sun and is usually determined by extrapolation of the dependence of intensity of radiation on the state of solar activity to the value which corresponds to zero area of spots and zero flare fields.

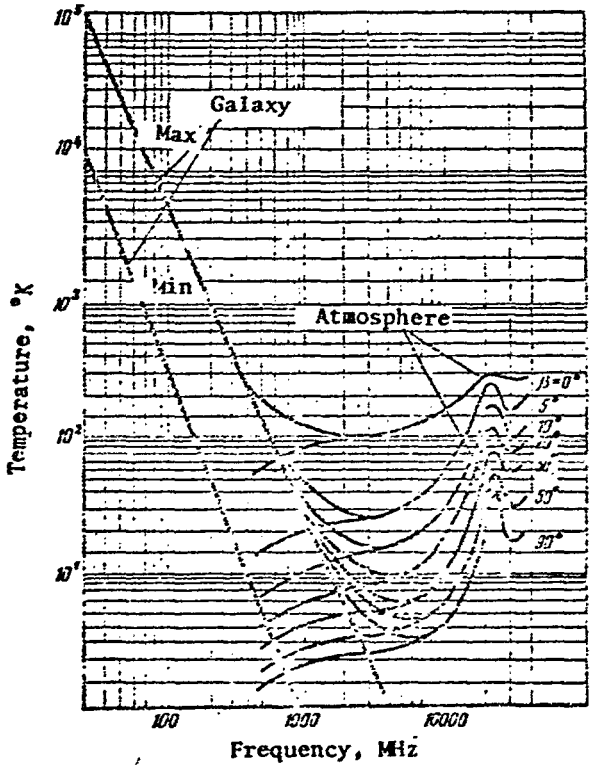


Figure 7.27. Equivalent Temperature of Antenna Caused by Cosmic Noise and Atmospheric Absorption

The spectrum of the thermal component is continuous, depends little on solar activity cycle and remains constant for a number of years. Its polarization is disordered.

The dependence of the equivalent temperature of radio radiation of the quiet sun on wavelength in the range studied is near linear and is rather well approximated by the formula [7.18, 7.19] /199

$$T_m \approx (5\text{-}6)10^3\lambda,$$

where λ is expressed in centimeters.

The intensity of radiation of the "perturbed" sun as a function of frequency may exceed the level of radiation of the quiet sun by many times (the meter wavelenghts, 10^2-10^3 times, at decimeter wavelengths -- by 10 times, at centimeter wavelengths -- by 2-3 times, at millimeter wavelengths -- by a few percent [7.18, 7.20]).

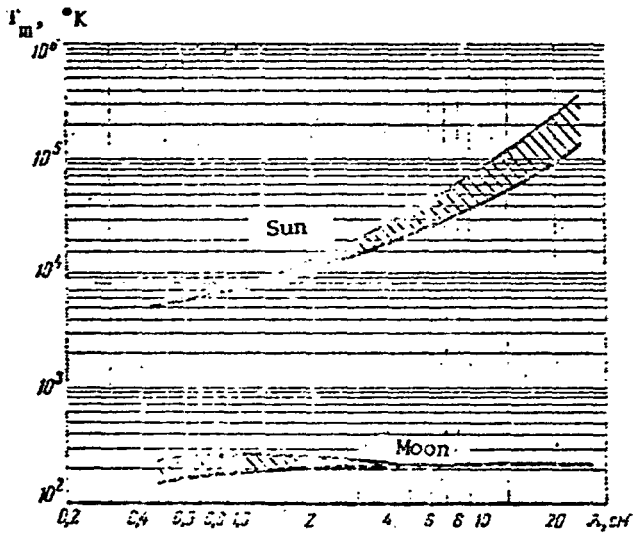


Figure 7.28. Equivalent Noise Temperature of Sun and Moon

Figure 7.28 shows the equivalent temperature of radiation of the sun as a function of wavelength. The upper curve corresponds to the years of maximum solar activity, the lower curve -- to the years of the minimum. As follows from the Figure, the sun is a rather powerful source of radio radiation and can fully disrupt

communications if it falls in the main lobe of the reception pattern of the ground station antenna. However, due to the small angular dimensions (visible diameter of the sun 0° 32'), if sharply directional antennas are used on the Earth the probability of this is very slight.

In the first approximation, this probability can be defined as the ratio of the surface occupied by the sun to the entire surface of the celestial sphere visible at angles $\beta \geq 5°$: /200

$$p = \frac{D^2}{(\pi - \beta)^2} \approx 8{,}5 \cdot 10^{-5}.$$

These discussions are based on the assumption that the ground station antenna can be directed toward any point in the sky. However, for all practical cases (for any orbit), the antenna is aimed only toward a limited area of the celestial sphere. In this case, the probability of "meeting" the sun may be still less.

Considering that interference with a space communications line is created when the direction from the ground station toward the satellite corresponds to the direction from the ground station toward the sun, and knowing the reception conditions (width of reception pattern of ground station antenna, orbital parameters of the satellite, signal/noise ratio at the input of the receiver), we can calculate both the maximum and total time of disruption of communications due to the influence of the radiation of the sun.

TABLE 7.6

H, km	v_{sat}, deg/hr	v_s, deg/hr	T_{max}, min
10400	~45	15	~1.4
13900	~30	15	~1.9
20300	~15	15	~2.8
35800	0	15	~5.6

Thus, for example for a satellite in an equatorial orbit, the maximum time of passage through the danger zone is

$$T_{max} = \left| \frac{\Delta\phi + D}{v_s \pm v_{sat}} \right|$$

where $\Delta\phi$ is the width of the reception pattern of the receiver antenna;

D represents the angular dimensions of the danger area of the sun (determined by reception conditions);

v_s and v_{sat} are the angular velocities of the sun and the satellite relative to the ground station respectively.

The plus sign in this formula relates to the case of motion of the satellite in an eastern direction, the minus sign -- in the western direction.

This formula ignores the angle between the trajectory of movement of the sun and of the satellite, i.e., the angle between v_s and v_{sat} at the moment it passes through the danger zone, due to its smallness.

For non-equatorial orbits, this angle must be considered.

Table 7.6 [7.21] shows the values of T_{max} for various altitudes of equatorial orbits of a satellite, calculated using this formula assuming that the satellite moves in the eastern direction and $D = 1.4^\circ >> \Delta\phi$.

It should be noted that the satellite passes through the center of the solar disc relatively rarely, and usually intersects it on lines displaced relative to the center. These lines move across the disc due to the change in inclination of the visible orbit of the sun. /201

Thus, Figure 7.29 [7.21], based on experimental data produced in October 1965 for the stationary Early Bird satellite, shows the displacement of its visible orbit relative to the center of the solar disc.

The abcissa shows the minutes (beginning at 14:09), while the ordinate shows the angular distance relative to the center of the solar disc.

As follows from Figure 7.29, the maximum duration of interference from the sun is observed on 9 October.

We can use this figure to determine the total duration of discuption of communications. To do this, we must draw a straight line parallel to the abcissa at the D/2 level, find the duration of interruption of communications for each day from the points of intersection of the corresponding curve and the line and add the results produced.

Thus, where $D = 1.4^\circ$, the duration of total disruption of communications for the month is about 15 minutes, for the year -- 30 minutes.

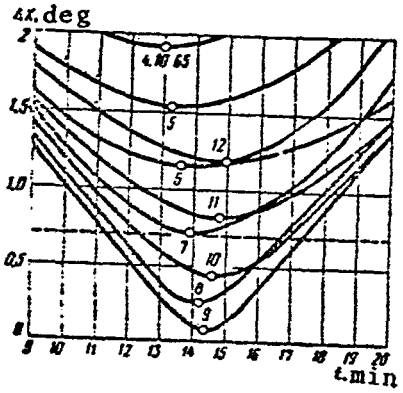


Figure 7.29. Trajectory of Motion of Satellite Relative to Center of Solar Disc (Early Bird, October 1965, antenna at Raisting)

As follows from [7.21], the duration of total disruption of communications calculated for a stationary orbit will not be exceeded for all other equatorial, as well as elliptical orbits of satellites, regardless of the geographic latitude of the ground station.

Thus, under the reception conditions selected (D = 1.4°), the probability that the sun will fall into the reception pattern of the receiver antenna for the most unfavorable months is

$$\frac{15}{30 \cdot 24 \cdot 60} = 3{,}5 \cdot 10^{-4}.$$

For D = 0.5°, this quantity will be $5 \cdot 10^{-5}$, i.e., will correspond to the results produced according to the approximate formula presented above.

The radiation of the moon and planets can create significant interference for space communications lines.

The equivalent radiation temperature of the moon at wavelengths shorter than 10 cm depends little on wavelength and phase of the moon; at longer wavelengths, this dependence disappears completely [7.22]. /202

Figure 7.28 shows the frequency dependence of the equivalent temperature of the moon, constructed on the basis of experimental data published in a number of works. The shaded area on the figure represents the possible fluctuations in radiation temperature of the moon as a function of its phase.

Thus, in the wavelength band from 3 cm up, the radiation temperature of the moon is almost independent of frequency and is equal to about 220°K; therefore, when working in this range, the rise in the equivalent noise temperature for an antenna with a gain of over 50 db, when aimed precisely at the moon, is approximately 220°K. The probability of this event, as for the sun, is not great, since the angular dimensions of the moon and the sun are the same.

Figure 7.30 shows the theoretical dependences of radiation intensity of the planets on wavelength [7.23]. The dotted lines

also show the experimental values of intensities for Jupiter and Venus.

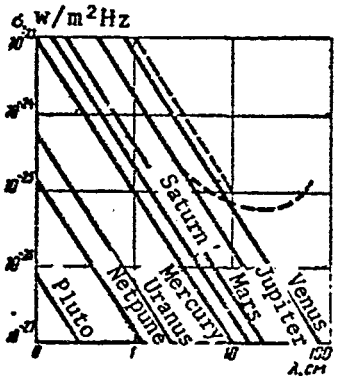


Figure 7.30. Intensity of Radiation of Planets of the Solar System

It is assumed that the divergence between theoretical and experimental values for thermal radiation of Jupiter at wavelengths of over 5 cm results from the fact that longer radio waves are generated by high energy electrons captured by the magnetic field of Jupiter.

Since 1946, discrete sources of radiation have been recorded with comparatively small angular dimensions. Their number has presently reached 2000, most of them identified with optically observed nebulae, as the results of more precise measurements of coordinates and angular dimensions of these sources have shown.

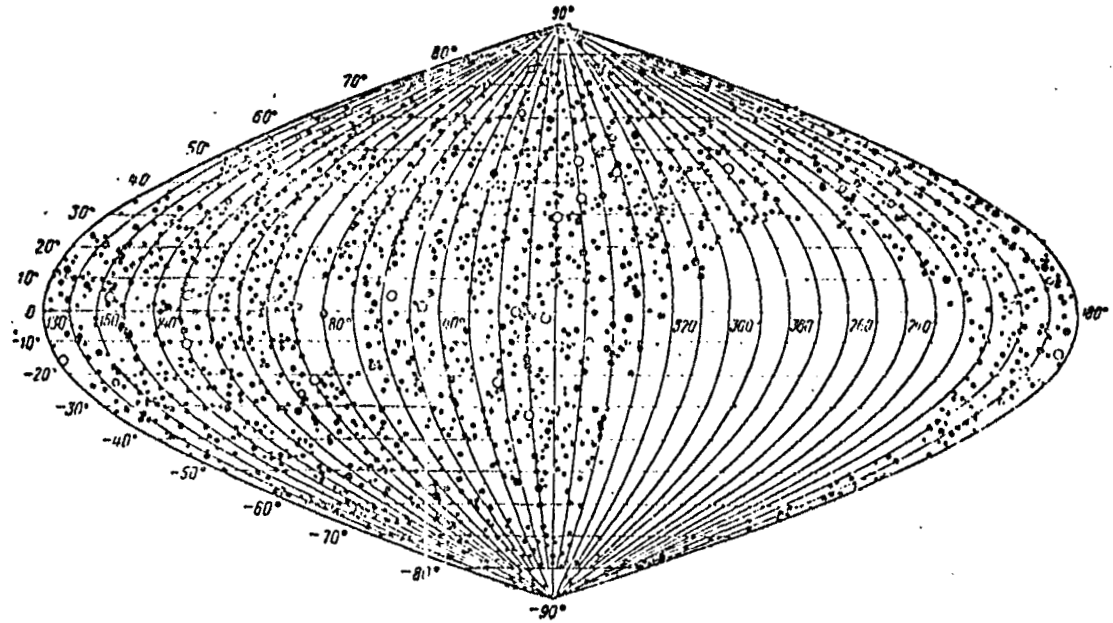
Some 20 of the most intensive, largest sources are located within ±5° of the galactic equator, while the remainder are distributed isotropically in space.

Figure 7.31 [7.24] shows a map of the distribution of 1936 sources visible from a single point (Cambridge, England). The white circles represent sources, the angular dimensions of which are over 20', the black circles -- sources with smaller angular dimensions. In both cases, the dimensions of the circles correspond approximately to the radiation flux. With a few exceptions, the radiation of discrete sources is not polarized. /205

Table 7.7 [7.25] presents data on the position, angular dimensions and radiation of the brightest of these. For extended sources of radiation, the table shows the maximum radiation flux density, which occurs when the solid angle of the antenna pattern is equal to or greater than the angular dimensions of the source.

As follows from this table, the dependence of radiation on wavelength for various sources differs. For some, the flux density increases with wavelength (Cassiopeia-A, Signus-A), while for others it remains approximately constant (Taurus-A, Omega nebula), while for still others it drops (Orion, Signus-X, Sagittarius).

Thus, the radiation of discrete sources creates interference in all wave bands. It is rather difficult to consider the influence of these sources, since there are a large number of them, distributed over the entire sky.



Diameter

>20'

<20'

Intensity

>81-10

41-80

21-40

0-20

Figure 7.31. Distribution of Discrete Sources in Galactic Coordinates

TABLE 7.7

Source	Flux Density W, 10^{-24} w·m^{-2}HZ^{-1} for Wavelength								Coordinates		
	3.2 cm	10 cm	20 cm	50 cm	70 cm	1 m	3 m	10 m	Right Ascension	Declination	Angular Dimensions
Cassiopeia-A	6	15	25	40	57	60	200	400	23:21	+58°30'	1.5'
Cygnus-A	-	8	12	20	40	35	110	200	19:58	+40°35'	1.5'
Taurus-A	7	8	10	13	13	16	18	18	05:31	+22°04'	6'
Virgo-A	-	2	2.5	3	5	5	12	-	12:28	+12°44'	5'
Centaurus-A	-	3	3	4.5	12	-	20	-	13:22	-42°46'	25'
Orion Nebula	3	4.5	4.5	-	3.5	1	-	-	05:33	-0.5°37'	-
Omega Nebula	7.5	7	8	-	-	-	-	-	18:17	-16°	-
Cygnus-X	-	85	68	34	-	27	3	-	20:20	-40:001'	10° x 2°
Sagittarius	-	400	350	240	-	-	300	-	17:43	-29°	12° x 2°
Nebula NGC 4628	-	-	-	-	-	-	350	-	16:51	-45°	10° x 6°

However, the number of powerful sources, as we have noted, is very low, and most of them create interference below the threshold of sensitivity of modern equipment.

The radiation of discrete sources is not polarized, so that reception on a polarized antenna can reduce it by 3 db.

Among the possible types of interference sources, we should turn particular attention to the radiation of the polar auroras, which may create significant interference during years of high solar activity for operations in polar regions. For example, in 1949 during a period of high solar activity, observations at 3 GHz recorded powerful auroral radiation [7.26]. The radiation power flux reached $2 \cdot 10^{-10}$ w/m^2·Hz.

We should also note the possible interference of radiation caused by lightening in the atmosphere. Although the maximum of radiation of these discharges is in the longwave band, the harmonics at higher frequencies may be significant.

§ 7.13. Determination of Equivalent Noise Temperature of Radiation of the Atmosphere

The radiation of the Earth's atmosphere results from the presence of the troposphere and the ionosphere and is thermal in nature. The primary sources of radiation in the atmosphere are oxygen, water vapor and precipitation. According to [7.4], the brightness temperature of the atmosphere can be determined from the following formula:

$$T_{ba} = T_{am} L_a, \tag{7.31}$$

where T_{am} is the mean temperature of the atmosphere; /206

L_a is a quantity defining the total absorption of radio waves in the atmosphere.

As calculations performed for the atmosphere model described in § 7.7 have shown, the mean temperature of the atmosphere for angles $\beta \geq 5°$ in this frequency band can be determined with satisfactory accuracy using the following formula:

$$T_{am} = T_0 - 32°K.$$

Considering this, the expression for the brightness temperature of the atmosphere, considering formula (7.20), can be reduced to the following form:

$$T_{ba} = (T_0 - 32°K)[(L'_{O_2} l_1 + L'_{H_2O} l_2) \operatorname{cosec} \beta + L_r(\beta)]. \quad (7.32)$$

Where $L_r(\beta) = 0$, i.e., where there is no precipitation, the brightness temperature of the radiation of the atmosphere is fully determined by its mean temperature and the coefficients of absorption of oxygen and water vapor at the surface of the Earth.

As was shown in § 7.7, losses of signal in a quiet atmosphere at 6 GHz do not exceed 0.5 db (where $\beta \geq 5°$). Substituting this value into the preceding formula, we produce

$$T_{ba} = (293°K - 32°K) \cdot 0.112 = 30°K.$$

Figure 7.27 shows the frequency dependence of brightness temperature of the atmosphere, calculated from the formulas presented for the entire range of angles β. Let us recall that these data relate to the quiet atmosphere, i.e., where $L_r = 0$.

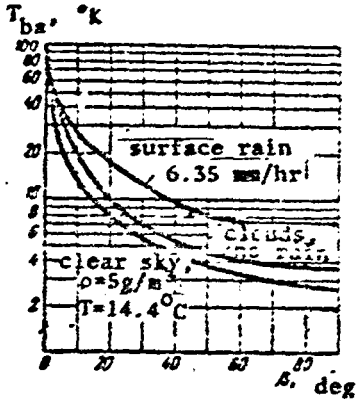


Figure 7.32. Change in Brightness Temperature of Atmospheric Radiation with Elevation Angle During Rainfall

The presence of precipitation can significantly influence the brightness temperature of the atmosphere. In principle, the influence of precipitation can be evaluated using the same formula (7.30), in which the value of L_r produced in § 7.7 should be used. The results of these calculations, as practice has shown, agree rather well with experimental data.

Figure 7.32 presents the results of measurement of the brightness temperature of the atmosphere at 6 GHz, performed in New Jersey, USA [7.3].

Figure 7.33 shows the theoretical results of calculation of brightness temperature at various frequencies for $\beta = 90°$ [7.27] with various intensities of precipitation.

It follows from Figure 7.32 and 7.33 that with intensive precipitation, the equivalent noise temperature at the antenna resulting from the radiation of the atmosphere may rise to (10-20)°K, even with the antenna pointed at the zenith. /207

Figure 7.32 also indicates that lower rain intensity in some cases yields a greater brightness temperature than rain at higher intensity. This phenomena is explained by the difference in thickness of the equivalent layer occupied by the precipitation.

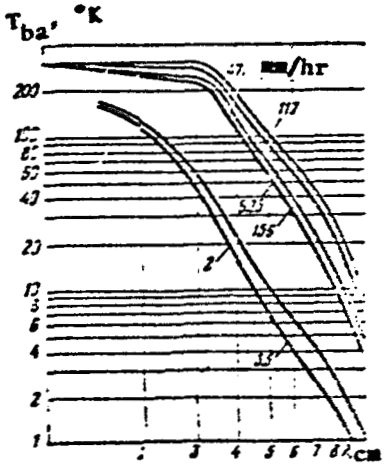


Figure 7.33. Change in Brightness Temperature of Atmospheric Radiation as a Function of Wave Length for Various Intensities of Precipitation as Rain

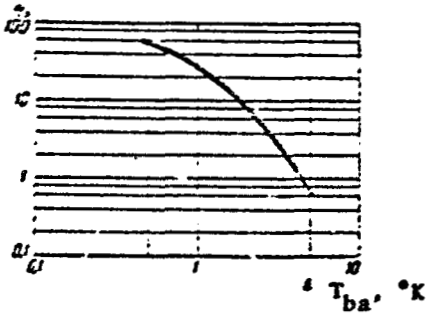


Figure 7.34. Statistical Characteristics of Slowly Changing Component of Brightness Temperature of the Atmosphere

Thus, the equivalent noise temperature of the atmosphere may increase significantly during heavy rain. However, as follows from Figure 7.32 and 7.33, for a rain intensity $i = 2$ mm/hr, the probability of exceeding which is less than 0.1% (see Figure 7.18), the brightness temperature of the atmosphere is not over 8°K where $\beta = 90°$ and 40°K where $\beta = 5°$, i.e., increases by 10°K in comparison to the quiet atmosphere.

We should also note that, in addition to the great fluctuations in noise temperature of the atmosphere during rainfall, slight changes in temperature are observed even when there is no precipitation.

It is assumed that these fluctuations in temperature result from passage of the antenna beam through clouds and changes in density of water vapor along the line of direct visibility in the direction of the antenna.

Figure 7.34 [7.28] shows a curve of the distribution of atmospheric temperature fluctuations constructed from the results of observations at 8 GHz, indicating that during a time not /208

exceeding 0.1% of the total time, the temperature may increase by 10°K even when there is no rain.

All of the quantitative estimates of components of the atmospheric brightness temperature presented above relate essentially to its lower portion -- the troposphere, since the influence of the troposphere is most significant in the frequency band in question.

The radiation of the ionosphere results from thermal radiation of ionized gas, and according to [7.4]

$$T_{bi} = T_{im} L_i(\beta),$$

where T_{im} = 1000-2000 °K is the mean kinetic temperature of electrons,

L_i is the total absorption of radio waves in the ionosphere.

As was indicated in § 7.7, absorption in the ionosphere is inversely proportional to the square of frequency; therefore, the value of T_{bi} in this frequency band is not over 1° and can be ignored.

§ 7.14. Determination of Temperature of Radiation of the Earth's Surface

According to [7.29], the brightness temperature of the level surface of the Earth is determined by the following expression:

$$T_{bg} = T_{0g}(1 - r_g^2), \tag{7.33}$$

where T_{0g} is the kinetic temperature of the Earth;

r_g is the reflection factor of electromagnetic energy from the surface of the Earth.

The reflection factors are determined by the following equations:

-- for vertical polarization

$$r_{gvert} = \frac{\varepsilon \sin\beta - \sqrt{\varepsilon - \cos^2\beta}}{\varepsilon \sin\beta + \sqrt{\varepsilon - \cos^2\beta}},$$

-- for horizontal polarization

$$r_{ghor} = \frac{\sin\beta - \sqrt{\varepsilon - \cos^2\beta}}{\sin\beta + \sqrt{\varepsilon - \cos^2\beta}}$$

where $\varepsilon = \frac{\varepsilon_g}{\varepsilon_0} - i\frac{\sigma_g}{\varepsilon_0 \omega}$ is the complex permittivity;

ε_g is the permittivity of the Earth;

ε_0 is the permittivity of the air;

σ_g is the conductivity of the Earth;

ω is the frequency.

For frequencies in the space communications range, the complex permittivity of the Earth can be taken as /209

$$\varepsilon \approx \frac{\varepsilon_g}{\varepsilon_0} = \varepsilon_r$$

(with the exception of large ocean surfaces).

Figure 7.35 [7.29] presents values of the reflection factor for power as a function of angle β for vertical and horizontal polarizations and the corresponding values of brightness temperature of the radiation of the Earth's surface.

The curves are constructed for the following conditions:

1) $\varepsilon_r = 4$, $\sigma = 0.0001$ mho/m (very dry ground);

2) $\varepsilon_r = 10$, $\sigma = 0.0001$ mho/m (very dry ground);

3) $\varepsilon_r = 30$, $\sigma = 0.001$ mho/m (moist ground);

4) $\varepsilon_r = 80$, $\sigma = 0.001$ mho/m;

5) $\varepsilon_r = 80$, $\sigma = 4$ mho/m, $f = 3$ GHz (ocean surface);

6) $\varepsilon_r = 80$, $\sigma = 4$ mho/m, $f = 1$ GHz (ocean surface).

In calculating the radiation temperature, the kinetic temperature of the Earth T_{0g} was assumed to be 290°K.

The calculations involved were for the level surface of the Earth and used the Fresnel formula for reflection factor. The concept of the "level surface" is applicable to the centimeter wave band only for a quiet ocean surface, paved surfaces, etc. In practice, we most frequently encounter uneven or, as they are also called, rough surfaces.

The radiation of the terrestrial surface in this case is determined by two components: the mirror component and the scattered or diffuse component. This general case is very difficult to calculate, and is therefore better studied experimentally.

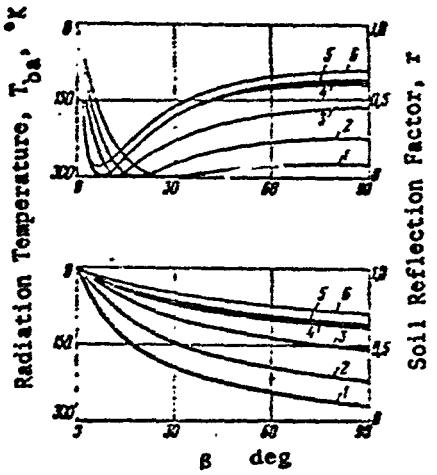


Figure 7.35. Power Reflection Factors as Functions of Elevation or Vertical (Upper) and Horizontal (Lower) Polarizations

With purely diffuse radiation, the brightness temperature of the radiation of the surface of the Earth is equal to its thermodynamic temperature (T_{0g} = = 290°K). Purely diffuse scattering is characteristic for surfaces covered with forests, underbrush or dense grass. In practice, the case of purely diffuse scattering is most frequently not realized. /210

The equivalent noise temperature of the antenna resulting from reflection of atmospheric radiation from the surface of the Earth can be calculated by the formula

$$T_{ba\text{-}g} = T_{ba} r_g^2. \qquad (7.34)$$

For an ideal reflecting surface, $r_g = 1$ and $T_{ba\text{-}g} = T_{ba}$. In practice, $T_{ba\text{-}g} < T_{ba}$, particularly for continental surfaces.

§ 7.15. Other Components of Antenna Noise

Of the possible sources of noise, we should also note the actual thermal radiation of the antenna itself, resulting from ohmic losses in its elements.

The equivalent noise temperature of the antenna resulting from internal thermal radiation can be estimated from the formula [7.4]

$$T_{nA} = T_0 L, \qquad (7.35)$$

where L are the losses resulting from absorption in the reflector of the antenna;

T_0 is the absolute temperature of the material of the antenna.

The results of calculations and experimental measurements have shown that antenna reflectors make a negligible contribution to the total noise temperature of the antenna.

As an example, Table 7.8 [7.4] shows the calculated values of equivalent noise temperature of the natural radiation of the antenna reflector for a steel reflector where T_0 = 290°K.

TABLE 7.8

λ, см	200	100	10	3	1	0,5
T_{nA}, °K	0,012	0,018	0,06	0,09	0,18	0,3

In certain cases, in order to protect the structure of a ground station antenna from the influence of external action (solar radiation, wind, precipitation, etc.), the antenna is covered by a radome.

The presence of a semitransparent (for radio signals) radome over the antenna results in additional losses of energy of the arriving wave and an increase in the equivalent noise temperature of the antenna, both due to the thermal radiation of the radome itself, and due to the reception of the radiation from the surface reflected from the radome.

As the results of theoretical and experimental studies have shown, the least increment in equivalent noise temperature of antennas is provided by radomes with wall thicknesses much less than the wavelength. /211

Precipitation in the form of rain or wet snow has a significant influence on the properties of a radome. When precipitation falls on the surface of the radome, a water film is formed, the presence of which increases losses and the noise temperature of the radome.

As an example, Figure 7.36 [7.30] presents the results of measurements at 417 GHz, showing the equivalent noise temperature adjusted to the input of the receiver (solid curves relate to natural thermal radiation, dotted curves -- to reflected radiation from the Earth's surface) with various elevations for a radome over a horn-parabolic antenna installed at Andover (USA). A special material (polystyrene fiber) was used for the radome, with high transparency and heat resistance. The wall thickness of the radome is 1.74 mm.

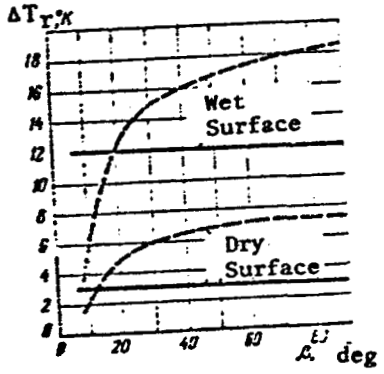


Figure 7.36. Increase in Equivalent Noise Temperature of Receiver System Resulting from Presence of a Radome : ——, Natural Thermal Radiation of Radome; ----, Terrestrial Radiation Reflected from the Radome

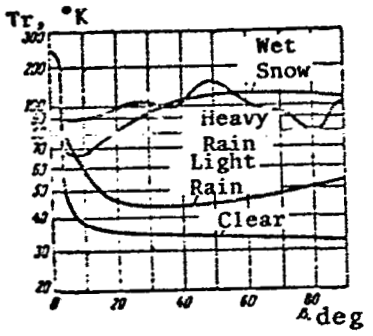


Figure 7.37. Changes in Equivalent Noise Temperature of Receiver System of Station at Andover as Functions of Elevation Under Various Weather Conditions

As the figure indicates, the equivalent noise temperature of the receiver resulting from reflection of the radiation of the terrestrial surface from the radome, in contrast to the temperature of the natural thermal radiation of the radome, depends on the orientation of the antenna aperture. Its value increases with increasing elevation. Furthermore, a significant increase in the noise temperature of the receiver occurs during precipitation.

The data presented agree /212 rather well with data from measurements of noise temperatures at the same frequency for a corner-reflector antenna radome used at Plémere-Bodue (France).

Figure 7.37 shows the results of measurement of noise temperature of the receiver at Andover at f = 4.17 GHz for various weather conditions. The significant increase in temperature at high elevations is explained by the accumulation of snow and rain water on the surface of the radome.

Thus, the presence of a radome increases the equivalent noise temperature of an antenna, which must be considered when low-noise devices are used in the ground station.

REFERENCES

7.1. Ayzenberg, G. Z., Korotkovolnovyye antenny [Shortwave Antennas], Svyaz'izdat Press, 1964.

7.2. Kalinin, A. I., Raschet trass radioreleynykh liniy [Calculation of Paths of Radio Relay Lines], Svyaz' Press, 1964.

7.3. Hogg, D., R. A. Semplak, "The Effect of Rain and Water Vapor on Sky Noise at Centimeter Wavelengths," Bell. Syst. Techn. I. 1961, Sept., V. 40, p. 1331-1348.

7.4. Tseytlin, N. M., Primeneniye metodov radioastronomii v antennoy tekhnike [Application of Radio Astronomy Methods in Antenna Technology], Sovetskoye Radio Press, 1966.

7.5. Arenberg, A. G., Rasprostraneniye detsimetrovykh i millimetrovykh voln. [Propagation of Decimeter and Millimeter Waves], Sovetskoye Radio Press, 1957.

7.6. ICCR Document IV/41-E, 19 January 1962.

7.7. Al'pert, Ya. L., V. L., Ginzburg, B. L. Faynberg, Rasprostraneniye radiovoln [Propagation of Radio Waves], Gostekhizdat Press, 1952.

7.8. Petrovich, N. T., Ye. F. Kamnev, Voprosy kosmicheskoy radiosvyazi [Space Communications Problems], Sovetskoye Radio Press, 1965.

7.9. Aarons, I., W. R. Barron, I. P. Castelli, "SHF Radio Astronomy Measurements," Zarubezhnaya radioelektronika, 1958, No. 10.

7.10. ICCR Recommendation No. 369, Geneva, 1963.

7.11. Deckett, M., TRW Internal Memo, May 1962.

7.12. Voronin, A. L., "Polarization Phenomena in the Magnetically Active Ionosphere," Geomagnetizm i aeronomiya [Geomagnetism and Aeronomy]. Vol. IV. 1964. No. 3. Nauka Press, Acad. Sci. USSR.

7.13. Voronin, A. L., "Refraction of Ultrashort Radio Waves in the Ionosphere," Geomagnetizm i aeronomiya [Geomagnetism and Aeronomy], Vol. IV, 1964, No. 3, Nauka Press, Acad. Sci. USSR.

7.14. Shubarin, Yu. V., Antenny sverkhvysokikh chastot [Super High Frequency Antennas], Khar'kov University Press, 1960.

7.15. Staras, H., "The Propagation of Wide-Band Signals Through the Ionosphere," Proc. IRE, 1961, Vol. 49, No. 7, p 1211.

7.16. Kolosov, A. A., Rezonansnyye sistemy i rezonansnyye usiliteli [Resonant Systems and Resonant Amplifiers], Svyaz'izdat Press, 1949.

7.17. Roman, N. G., B. S. Iaplee, "Radio Sources and the Milky Way at 440 MC," Proc IRE, 1958, Vol. 46, No. 1, p. 199.

7.18. Zheleznyakov, V. V., "Radio Radiation of the Sun and the Planets," Uspekhi fizicheskikh nauk, Vol. 64, 1958, No. 1, p. 113.

7.19. Du Lep-Yao, A. N. Malakhov, et al., "Observation of the Annular Eclipse of the Sun of 19 April 1958 at Wavelengths of 1.63, 3.2 and 10 cm," Izvestiya vysshikh uchebnykh zavedeniy, Radiofizika, Vol. 2, 1959, p. 154.

7.20. Dodson, H. W., "Studies at the McMath-Hulbert Observatory of Radio Frequency Radiation at the Time of Solar Flares," Proc. IRE, 1958, Vol. 46, No. 1, p. 149.
7.21. Erler, G., M. Schonfeld, "Disturbances of Communications Between Satellites Due to the Sun," Nachrichtentechn Z, 1966, No. 11, pp. 653-658.
7.22. Troitskiy, V. S., V. D. Krutikov, Uspekhi fizicheskikh Nauk, XXXI, No. 4, 1963.
7.23. Mayer Cornell H., "The Temperatures of the Planets," Suent. Amer, 1961. Vol. 204, No. 5, pp. 58-65.
7.24. Shklovskiy, I. S., Kosmicheskoye radioizlucheniye [Cosmic Radiation], Gostekhizdat Press, 1956.
7.25. Yevdokimov, M. A., "Cosmic Radiation," Zarubezhnaya Radioelektronika, 1960, No. 7.
7.26. Little, G. C., et al., "Review of Ionospheric Effects on VHF and UFH," Proc. IRE, 1956, Aug., Vol. 44, No. 8, p. 992.
7.27. Hogg, D. C., P. A. Semplak, "Estimate of Temperature of Atmospheric Noise in the Super High Frequency Band During Rain," Proc. IRE, Vol. 51, 1963, No. 3.
7.28. Orhaug, T. P., Proceedings of the IEEE, Vol. 51, No. 3, March 1963.
7.29. Croom, D. L., "Naturally Occurring Thermal Radiation in the Range 1-10 Gc/s," AIEE, 1964, Vol. 11, No. 5, pp. 967-980.
7.30. Tocquec, Y., "Theoretical Studies of Limitations Resulting from Characteristics of Cold Antennas with High Gain," 14th International Astronaut. Congr. Paris, 1963, Vol. 2, Paris-Warsaw, 1965.

CHAPTER 8. MULTIPLE ACCESS OPERATION AND METHODS OF MULTIPLEXING OF RADIO SIGNALS

§ 8.1. Statement of Problem and Selection of Criteria /214

The condition of multistation operation, i.e., simultaneous operation of a large number of ground stations through a single relay satellite, places specific and serious requirements on the methods of transmission of signals in satellite communications systems.

The task of planning consists in the selection of a method of modulation and a method of multiplexing of radio signals in the relay trunk such that the transmission capabilities of the relay station (output power and band of frequencies) are used most completely, while the level of mutual interference between signals is minimal, and has the minimum effect on reception conditions of each of the signals relayed. It must be considered in this case that the relay station itself can introduce distortions and create noise; consequently, ground station signals should be resistant to these noises.

Thus, the problem is to find a set of signals which are independent or near independent, having the greatest resistance to the interference and distortion characteristic for the relay station used. Three main methods of formation of a set of orthogonal signals are known: frequency multiplexing, time multiplexing and shape multiplexing.

In the first case, each signal is assigned an individual frequency band and the spectrum of signals does not overlap. In this case, all of the signals multiplexed have identical form and all are relayed simultaneously and continually.

In the second method, independence of the signals is achieved by their time separation; the form of the signals and their spectra may correspond completely in this case.

Finally, in the third method all of the signals are relayed simultaneously and occupy a common frequency band. However, the signals differ in form, allowing them to be separated at the reception point. The methods may also be combined. For example, /215
the combination of frequency and time multiplexing has led to the development of today's discrete-address systems with simple signals, and the discrete-address systems with complex noise-like signals result from simultaneous multiplexing by frequency, time and form.

In each of these cases, certain specific noises arise when the signals are relayed. For example, with frequency division multiplexing, the nonlinearity of the amplitude characteristic of the relay station causes combination products to appear, with time division multiplexing the neighboring signals may overlap, while waveshape multiplexing characteristically shows interference resulting from the nonorthogonality of the waveforms. Therefore, in each case we must select a waveform (or modulation method) minimizing the influence of these types of interference.

To evaluate various methods of transmission, we must establish certain criteria which most completely reflect the essence of the task to be performed. The known criteria used to evaluate method of modulation, such as the effectiveness of the utilization of signal power (β effectiveness) or the effectiveness of utilization of the frequency band (α effectiveness) are insufficient in this case, since they fail to consider the development of mutual interference and, consequently, do not encompass the entire problem.

The best and most complete criterion is the criterion of mutualization of the throughput capacity of the relay station during multistation operation, since it not only reflects all of the energy relationships between signals and noise upon transmission of a large number of signals through a single line, but also provides broad possibilities for economic evaluation of the entire satellite communications system.

The essence of this criterion is as follows. Given a relay station with output power P, frequency band W and operating time T. The throughput capacity of this relay station is equal to C, i.e., it can transmit M simplex telephone channels between two ground stations (i.e., on a single carrier).

With multistation operation and any given method of multiplexing, this trunk is divided into n subtrunks, each of which is assigned to one ground station. The parameters of a subtrunk P_i, W_i and T_i are such that its throughput capacity is equal to C_i, i.e., it can transmit only N' telephone channels.

Accordingly, the throughput capacity of the trunk for multistation operation is equal to

$$C_\Sigma = \sum_{i}^{n} C_i \quad \text{or} \quad N_\Sigma = nN'.$$

The criterion which we will use is the ratio /216

$$Z = \frac{C_\Sigma}{C} \quad \text{or} \quad Z = \frac{N_\Sigma}{N}$$

characterizing the degree of utilization of the throughput capacity of the relay station in multistation operation. Since the throughput capacity of each subtrunk depends on its power parameters, determined by the method of modulation used, while C_Σ, naturally, depends on the number of subtrunks n, the expression for coefficient Z in this case can be written as follows:

$$Z(n)=\frac{\sum_{i}^{n} C_i(P_i, W_i, T_i, B_i)}{C(P, W, T, B)}, \tag{8.1}$$

where B and B_i characterize the gain in signal/noise ratio provided by the modulation method selected.

It follows from (8.1) that quantity Z is a function of the number of signals multiplexed (i.e., the number of subtrunks). It follows from general heuristic considerations that this dependence cannot be an increasing function of n; in the best case, the throughput capacity of the relay station in the multistation mode remains constant; in the worst case, the transition to the multistation mode will be accompanied by a loss of throughput capacity, this loss being the natural price of the operational advantages and conveniences provided by multistation access.

Thus, from the standpoint of the criterion used, the problem of evaluation of methods of signal transmission consists in determination of the functional Z(n) and selection of a method providing the highest value of Z with a fixed number n of stations in a system.

According to this, let us study a number of methods of modulation and multiplexing and establish how completely they correspond to the requirements placed on them. Our primary attention will be given to the peculiarities and effects arising upon multiplexing of signals, since all problems concerning methods of modulation in general have been studied in the basic work [8.1] and, as applicable to satellite communications, are described in [8.2] and [8.3].

§ 8.2. Frequency Multiplexing. Nonlinear Noise and AM-PM Transition Interference

The simplest and most widely used method is frequency division multiplexing, in which each ground station transmits its signal in a sector of the frequency spectrum set aside for it (Figure 8.1).

Guard frequency intervals Δf_g are set aside between the signals of the ground station, allowing signals to be separated at the reception point with the required accuracy.

Thus, with frequency multiplexing, several signals pass through the relay station at the same time. Studies of the parameters of such a multistation signal have shown that a good model for it is the sum of sinusoidal signals with random phase. The distribution function of this signal is /217

$$w(x)=\frac{1}{2n\alpha}\left[1+2\sum_{r=1}^{\infty}\cos\frac{\pi r x}{n\alpha}J_0^n\left(\frac{r\pi}{n}\right)\right] \text{ where } |x|<n\alpha,$$

where a is the amplitude of a partial signal;
n is the number of signals multiplexed;
J_0 is a zero order Bessel function.

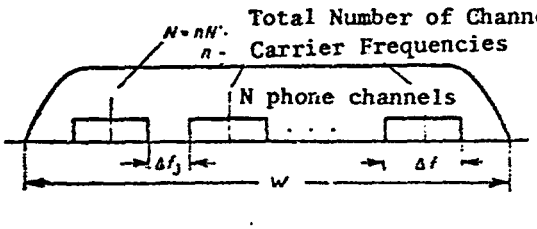


Figure 8.1. Frequency Division Multiplexing of Ground Station Signals

The peak factor of a multistation signal, defined as the ratio of the peak power, not exceeded with probability p(λ), to the mean power n of harmonic oscillations, is

$$v^2=\frac{2\lambda^2\alpha^2}{n\alpha^2}=\frac{2\lambda^2}{n}, \qquad (8.2)$$

where λ is the peak level of the multistation signal, the probability of exceeding which is defined by the equation

$$p\{|\varepsilon|<\lambda\alpha\}=\frac{1}{n}+\frac{2}{\pi}\sum_{r=1}^{\infty}\frac{1}{r}J_0^n\left(\frac{r\pi}{n}\right)\sin\frac{\pi r\lambda}{n}. \qquad (8.3)$$

Figure 8.2 shows values of the peak factor of a multistation signal, calculated using formulas (8.2) and (8.3), for various values of n. It follows from this figure that with probability p(λ) = 0.999, the peak level of the multistation signal exceeds its mean level by not more than 10-11 db. /218

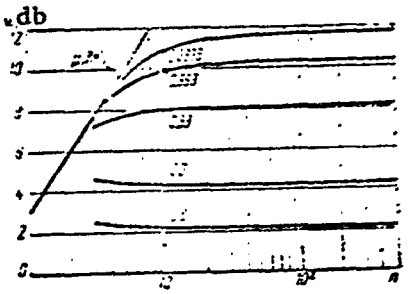


Figure 8.2. Peak Factor of Multistation Signal with Frequency Division Multiplexing

The transmission of this complex signal through the relay station results in the appearance of a number of factors specific for frequency division multiplexing. They include: reduction of the mean output power of the relay station, mutual suppression of signals, appearance of crosstalk interference due to nonlinearity of the amplitude characteristic of the relay station and, finally, appearance of cross talk interference due to the presence in the relay station of elements which convert the amplitude modulation of the signal to phase modulation (AM-PM conversion). Let us study these effects briefly.

Figure 8.3 shows the amplitude characteristic of the relay station, approximated by a broken line. Although the true characteristic differs from the characteristic presented, this idealization is convenient for calculations, and the results produced using it agree well with experimental data.

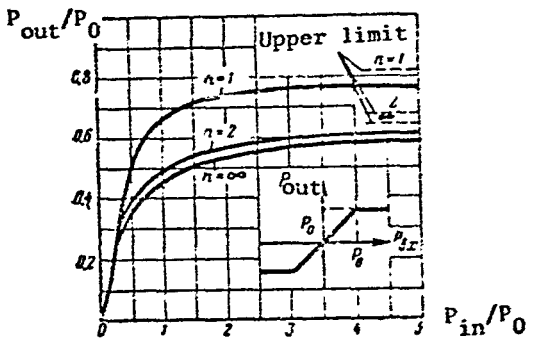


Figure 8.3. Determination of Losses of Output Power of Relay Station with Frequency Division Multiplexing

As is shown in [8.4], when this type of relay station amplifies one harmonic signal, its output power, normalized relative to saturation power P_0, is equal to

$$\frac{P_{1out}}{P_0} = \frac{2}{\pi^2}\left[\left(1-\frac{1}{2\rho}\right)^{1/2} + (2\rho)^{1/2}\sin^{-1}\left(\frac{1}{2\rho}\right)^{1/2}\right]^2, \qquad (8.4)$$

where

$$\rho = \frac{P_{1in}}{P_0},$$

and with high power of the input signal, corresponding to the saturation area ($P_{1in} >> P_0$)

$$P_{1out} \approx \frac{8}{\pi^2} \approx 0{,}81. \tag{8.5}$$

When two equal harmonic signals are amplified, their summary output power is

$$\frac{P_{1out} + P_{2out}}{P_0} = \rho \left[1 + \frac{4}{\pi^2} F(\rho)\right]^{-2} = \left(\frac{8}{\pi^2}\right)^2 \approx 0{,}65 \quad \text{where } \rho \to \infty. \tag{8.6}$$

When a large number of signals is amplified, when their sum has the properties of white noise, the output power of the sum of these signals /219

$$\frac{\sum_{i=1}^{n} P_{i\,out}}{P_0} = \rho \operatorname{erf}^2 \left(\frac{1}{2\rho}\right)^{1/2} = \frac{2}{\pi} \approx 0{,}64 \quad \text{where } \rho \to \infty. \tag{8.7}$$

The results of calculations using formulas (8.4)-(8.7) are presented on Figure 8.3, from which we can see that the output power of a relay station in the multistation mode is less than when a single signal is amplified by 1-1.5 db. Actually, this is explained by the fact that when operating in the area of saturation (or near it), crosstalk products are formed, and part of the power of the relay station is expended on these products.

As we know, the transmission of signals of various levels through a nonlinear system is accompanied by the suppression of the weak signal by the strong signal. The suppression factor is defined as the ratio of output powers of the signals to their input powers, i.e.,

$$K_s = \frac{P_{1out}/P_{2out}}{P_{1in}/P_{2in}}.$$

With sinusoidal signals and a system amplitude characteristic approximating the characteristic of an ideal limiter, the following relationships are correct:

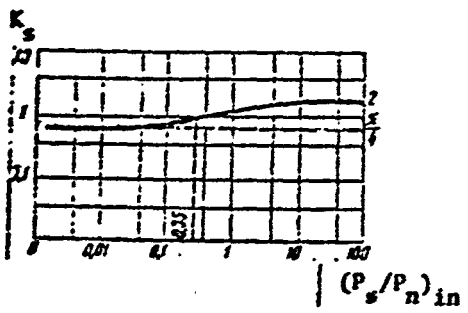


Figure 8.4. Signal/Noise Ratio at Output of Limiter

$K_s \to 4$ where $P_{1in} \gg P_{2in}$,

$K_s = 1$ where $P_{1in} \approx P_{2in}$,

$K_s \to 0.25$ where $P_{1in} \ll P_{2in}$.

A similar effect is observed with joint transmission of a signal and noise through the system, although the numerical values of K_s are different. The dependence of K_s on the ratio of signal power to noise power at the input of the system is shown in Figure 8.4, from which we see that where $P_s \gg P_n$, the value of $K_s \to 2$, while where $P_s \ll P_n$, on the other hand, $K_s \to \pi/4$.

Thus, in the worst case the suppression factor of the useful signal by the interfering signal reaches 6 db (with harmonic interference), never exceeding 1 db for suppression by noise; the greatest suppression of noise by a harmonic signal is 3 db.

The suppression of weak signals forces us to use regulation of ground station power during frequency division multiplexing in order to be sure that the power levels of all signals at the input of the relay station are approximately the same. /22

The most difficult problem associated with frequency division multiplexing is reduction of the level of crosstalk interference caused by nonlinearity of the system and the AM-PM conversion.

In order to estimate the products of crosstalk modulation resulting from nonlinearity of the amplitude characteristic of the relay station we can utilize the results of [8.5] produced for the limiting case when the number of signals multiplexed is great and the energy spectrum of the entire set of modulated carriers can be considered even (Figure 8.5).

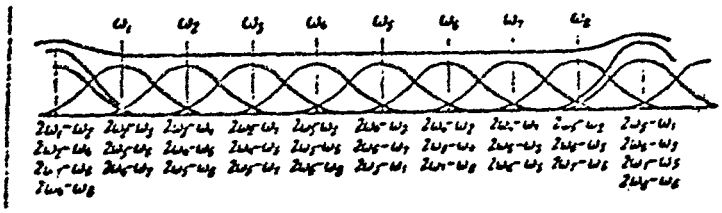


Figure 8.5. Spectrum of Third Order Nonlinearity Products

The main share of crosstalk noise results from third and fifth order combinations. The power of crosstalk noise, measured in the frequency band of one signal is

$$P_{nn} = \frac{\Delta f}{W} 12e^{-2b_g} 6b_m, \text{ mw}, \qquad (8.8)$$

where Δf is the band of frequencies of the ground station receiver;

W is the band of frequencies of the relay station;

b_g is the attenuation of nonlinearity in the system at the third harmonic, in nepers, with a measurement signal with a level in nepers equal to the level of one carrier;

b_m is the ratio of the mean level of the multistation signal to the mean level of one signal, nep.

With a small number of multiplexed signals, the power of third order combination interference (i.e., interference such as $2\omega_1-\omega_2$ or $2\omega_2-\omega_1$) according to [8.4] is equal to

$$P_{nn} = \frac{4a}{3\pi^2}\int_0^{\pi}(1-u^2)^{3/2}\cos y\,dy, \qquad (8.9)$$

where

$$u=\rho^{-1/2}+\cos y;\quad \theta=\cos^{-1}(1-\rho^{-1/2}),$$

a is the amplitude of the signal multiplexed.

A significant reduction in nonlinear interference level can be achieved by using uneven placement of frequencies of ground station signals over the frequency band of the relay station. With optimal placement, most of the products will be located between signals and will not fall directly in the frequency band of a receiver. For example, placement of frequencies in a geometric progression can reduce noise power by 8 to 10 db. /221

The presence in the relay system of elements converting the amplitude modulation to phase modulation results in the appearance of two types of conversion interference: intelligible and unintelligible.

The appearance of intelligible interference results from the fact that the signal of each ground station, modulated, for

example, as to frequency, as it passes through a route with non-uniform frequency characteristics, is given parasitic AM according to the frequency modulation rule (transition of FM to AM). When several signals are amplified in a single line, at least one of which has AM with the frequency of useful signals, the summary signal will also be amplitude modulated, and the modulation components will include the useful signal frequency.

After passage through an element with AM-PM conversion (for example a traveling wave tube), the amplitude modulation is converted to phase modulation and, consequently, the phase of each signal includes intelligible interference with the modulation frequency of the neighboring station.

The power of this interference can be determined by the formula

$$P_{\text{in AM-PM}} = \left[\frac{k_{\text{AM-PM}} A F_M}{57.2(R+1)}\right]^2, \qquad (8.10)$$

where K_{AM-PM} is the AM-PM conversion factor, degrees/db;

A is the non-uniformity of the frequency characteristic of the line, db/MHz;

R is the ratio of amplitudes of useful and interfering signals;

F_M is the modulation frequency of the interfering signal.

In order to estimate the power of unintelligible interference of the AM-PM conversion, we can use the results produced in [8.4]. With a small number of multiplexed signals, the power of unintelligible interference, normalized relative to the signal, will be

$$P_{\text{unin AM-PM}} = \frac{J_1^2(2k_0)}{J_0^2(2k_0)+J_1^2(2k_0)} \approx 1{,}45 \cdot 10^{-3} k^2 \text{ AM-PM}. \qquad (8.11)$$

With a large number of signals

$$P_{\text{unin AM-PM}} \approx 2.2 \cdot 10^{-3} k^2_{\text{AM-PM}}. \qquad (8.12)$$

The power spectrum of the unintelligible conversion interference will be the same as the spectrum of the products of nonlinear conversion (Figure 8.5).

We must note that the degree of nonlinearity of the characteristic of the relay station and the value of k_{AM-PM} depend on the level of the input power, or more precisely on the ratio of input power to saturation power. Figure 8.6 shows a typical dependence of each of the components of the interference, calculated according to formulas (8.4)-(8.12), as well as their sum $P_{n\Sigma}$, on the input power for amplification of a large number of signals. The dotted line on this same figure shows the curve characterizing $P_{n\Sigma}$ for amplification of only two signals. /222

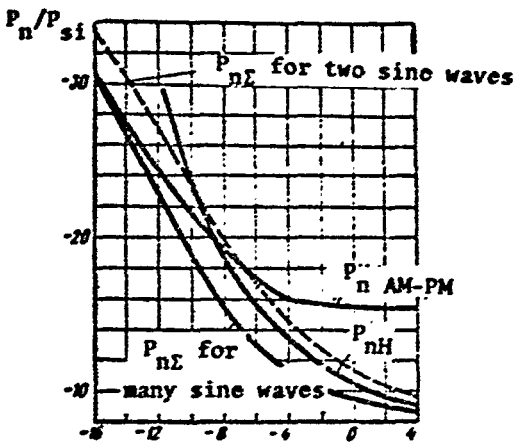


Figure 8.6. Power of Crosstalk Noise with Frequency Division Multiplexing

Comparison of these curves indicates that an increase in the number of signals relayed with the same mean power of the multistation signal increases the level of crosstalk interference by not over 4-5 db.

However, the absolute value of interference power depends essentially on the mean power of the signals at the input of the TWT and is decreased by more than 20 db by reducing the output power of the TWT relative to saturation power.

§ 8.3. Estimate of Effectiveness of Frequency Division Multiplexing

The presence of all of the effects which we have studied leads to the fact that the throughput capacity of a relay station with multiple access operation with division multiplexing lower than when information is transmitted over a single carrier.

Let us show this using the example of FM signals. In order to estimate the dependence Z(n), we must establish the relationship between the power of noise of various origins measured in a telephone channel and the number of signals multiplexed.

Precise solution of this problem is difficult, since it requires consideration of components not only of third and fifth, but also of higher orders; not only the power of these components, but also their spectral composition must be considered. This composition depends, in turn, on the nature of the product.

therefore, in the first approximation we can use the results of [8.6] for estimation of the throughput capacity of the relay station with frequency division multiplexing of FM signals, assuming that the power of crosstalk interference is independent of the number of signals multiplexed. As was indicated above, this assumption is realistically based. /223

The fixed noise power level in a telephone channel at the zero relative level point can be represented as two components:

$$P_{nk} = P_{nt} + P_{ni}$$

where P_{nt} is the power of thermal noise;

P_{ni} is the power of interference noise.

Although the conditions of the problem require only constancy of total noise power, for simplicity let us require also that as the number of stations changes, each of the components of noise power remains constant, i.e., P_{nt} = const and P_{ni} = const.

The first of these requirements can be fulfilled by proper selection of frequency deviation and power of the FM signal at the input of the ground station receiver, while the second is fulfilled by selection of the frequency band of the receiver and observation of necessary linearity of the amplitude characteristic of the relay station. In order to simplify the problem, let us introduce a certain idealization, assuming that the amplitude characteristic of the relay station is linear right up to the saturation point (see Figure 8.3), defining the peak power of the relay station, while the power of the products of the Am-FM conversion is independent of the number of carriers in the first approximation.

As we know [8.5], the signal/thermal noise ratio in the upper telephone channel of a multichannel FM receiver with frequency division multiplexing of channels in the group spectrum is equal to

$$\left(\frac{P_s}{P_{nt}}\right)_{out} = \left(\frac{P_s}{P_n}\right)_{in} B$$

and

$$B = \frac{1}{2}\frac{\Delta f}{\Delta F_k}\frac{\Delta f^2 dk}{f_k^2},$$

where $(P_s/P_n)_{in}$ is the signal/noise ratio at the input to the receiver in the band of the FM signal;

B is the gain in signal/noise ratio achieved by the FM receiver;

ΔF_k is the band of frequencies of one telephone channel;

Δf_{dk} is the peak deviation of frequency per channel of the multichannel message;

f_k is the mean frequency of the top telephone channel.

Let us represent the quantities included in this formula as functions of the number of stations in the system. In the first approximation, we can ignore the fact that the group spectrum of a multichannel telephone signal begins not at zero, but at a certain boundary frequency, and assume that /224

$$f_k \approx N' \Delta F_k a = \frac{N}{n} a \Delta F_k,$$

where N' is the number of telephone channels transmitted by one station;

a is a certain constant factor considering the presence of guard bands between neighboring channels with group frequency division multiplexing.

Further, it is known that the peak deviation of frequency per channel is related to the peak deviation of frequency of the entire multichannel message Δf_d by the following formula:

$$\Delta f_{dk} = \sqrt{2} \frac{\Delta f_d}{b_0},$$

where b_0 is the ratio of the peak level of the multichannel message to the level of the test signal in one telephone channel.

According to [8.5], the relative peak level b_0 of a group signal consisting of N' telephone channels with frequency division multiplexing can be determined by the formula

$$b_0(N') = -15 + 10 \lg N' + v$$

for $N' \geq 240$ telephone channels,
where v is the peak factor of the multichannel telephone message, determined with fixed probability according to the graphs of [8.5].

For a lower number of channels, the value of b_0 can be determined from graphs of $b_0 = f(N')$ presented in [8.5].

Finally, let us express by n the frequency band of a receiver. It follows from Figure 8.1 that with n signals and the band of frequencies of each receiver Δf, the number of guard bands of width Δf_g is n - 1, and the necessary bandwidth of the relay station

$$W = n\Delta f + (n-1)\Delta f_g.$$

In turn, we can assume that

$$\Delta f = p2\Delta f_d \text{ and } \Delta f_g = \Delta f(q_f - 1),$$

i.e., the band of frequencies of the receiver is p times greater than twice the deviation of frequency of the multichannel signal, and the guard band between signals of the station is proportional to the width of the band of frequencies of the receiver. The value of coefficient p is determined by the required level of transient noise introduced by the ground station receiver, and constancy of coefficient p in the first approximation (with high modulation indices) assures constancy of noise power.

From the last two expressions we produce /225

$$\Delta f_d \approx \frac{W}{2pnq_f}.$$

Let us return to the initial formula and determine the signal power at the input of the ground station receiver

$$P_{sin} = kT_{\Sigma g}\Delta f(P_s/P_n)_{in},$$

where k is Boltzman's constant;
$T_{\Sigma g}$ is the equivalent noise temperature of the ground station receiver.

Correspondingly, the mean power of the satellite transmitter required to transmit the signal of one station is

$$P_{si} = P_{sin} L_{\Sigma},$$

where L_{Σ} is the total losses of energy between the satellite transmitter and the ground station receiver.

In the system in question, n signals pass through the relay station simultaneously; the mean power necessary for transmission of all these signals

$$P_{sm} = P_{si} n,$$

while the peak power of the relay station

$$P_{speak} = P_{sm} \nu^2 = P_{sin} L_{\Sigma} n \nu^2,$$

where ν is the peak factor of the multistation signal.

Substituting the quantities defined by these formulas into a known formula, we produce a general expression relating all power parameters of the system:

$$P_{speak} = \frac{4 L_{\Sigma} k T_{z_3} \Delta F_k^3 a^2 \rho^2}{W^2} \quad \frac{P_s}{P_{nt}} \quad (N')^2 n b_0 (N') \nu^2 q_f^2.$$

Assuming P_{speak} = const and W = const, which is necessary in order to apply our criterion, we can use this formula to calculate the dependence of N' and N on the number of signals multiplexed. The values of peak factor of the multistation signal necessary in this case were presented earlier on Figure 8.2.

The results of calculations are presented in Figure 8.7, from which we see that a relay station capable of transmitting N = 600 telephone channels on one carrier (where n = 1), can transmit only N' = 8-12 telephone channels for each carrier with frequency division multiplexing of the signals of 10 stations, or, consequently, N = 80-120 channels (α = 0.13-0.2). The spread of these values results from the differences in width of guard bands, characterized by the factor q_f. Where n = 50-60, the throughput capacity of the relay station becomes minimal (only one telephone

channel can be transmitted for each carrier), amounting to only 8-10% of its initial value (Z = 0.08-0.1)[1]. /226

However, in this asymptotic mode, new capabilities appear in the system for increasing the throughput capacity by statistical multiplexing of communications channels [8.7].

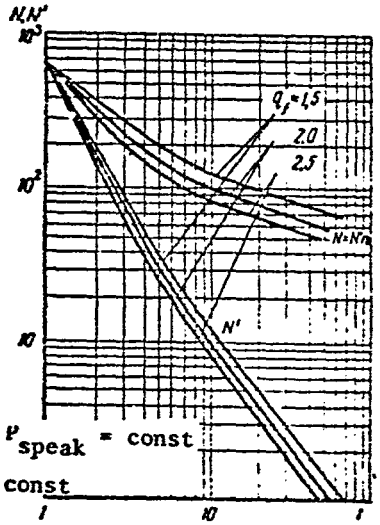


Figure 8.7. Throughput Capacity of a Relay Station with Frequency Division Multiplexing of FM Signals

As we know, a telephone channel is active only 25-30% of the time, during the time it is busy. This is explained by time losses for subscribers calling, the presence of pauses between phrases and words, and finally, during duplex operation, each subscriber is silent approximately half the time, listening to the person with whom he is conversing. If the carrier is switched off during pauses in the modulated signal, the statistical mean load on the relay station by ground station signals can be decreased, or the number of stations in the system can be increased with the same load.

In order to estimate the gain in throughput capacity provided by statistical multiplexing, let us calculate the peak factor of a multistation signal consisting of the sum of a random number of harmonic oscillations with random phase.

Using the full probability formula

$$p_{\xi_\nu}(x)=p\{\xi_\nu<x\}=\sum_{m=0}^{N}p(\nu=m)\,p(\xi_m<x)$$

and the formula for distribution of the sum of a fixed number of sine waves with random phase [8.3], we produce

[1]More precise consideration of all noise components and optimization of the parameters of the system considering interference could probably increase the throughput capacity of the relay station in comparison with that calculated above, but the gain could not be by more than a factor of 2 in the best case.

$$p_{\xi_n}(x) = \sum_{m=1}^{n} \binom{n}{m} p^m q^{n-m} \left[\frac{\lambda}{m} + \frac{2}{\pi} \sum_{r=1}^{\infty} \frac{1}{r} J_0^m \left(\frac{r\pi}{m} \right) \sin \frac{\pi r \lambda}{m} \right] \qquad (8.13)$$

The results of calculation by formula (8.13) are presented in Figures 8.8 and 8.9 in the form of the dependences of peak factor of a multistation signal with carrier suppression on the number of ground station signals n with various levels of activity of each channel, characterized by coefficient p_a, and various probabilities that this level will not be exceeded, characterized by coefficient p_c. /227

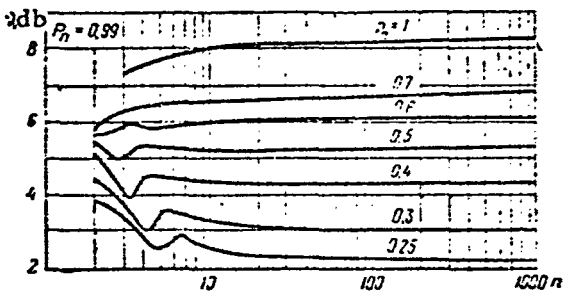


Figure 8.8. Peak Factor of Multistation Signal with Statistical Multiplexing with p_c = 0.99

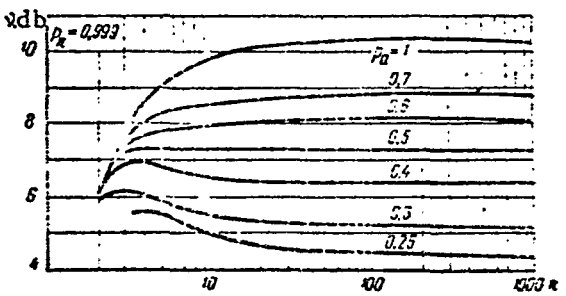


Figure 8.9. Peak Factor of Multistation Signal with Statistical Multiplexing for p_c = 0.999

It follows from the figures that with a fully active channel (p_a = 1), the peak factor of a multistation signal is approximately 8 db with probability p_c = 0.99, corresponding to the calculations performed earlier, the results of which are illustrated by Figure 8.2.

Where p_a = 0.25-0.3, the value of the peak factor decreases to 2-3 db with a sufficiently large number of carriers (n > 10).

Thus, suppression of carriers in this example either allows the peak power of the transmitter to be reduced by 5-6 db, or allows the number of signals relayed to be increased.

This latter path is most expedient. The gain in throughput capacity of a relay station is characterized by coefficient γ, the ratio /228

$$\gamma = \frac{\text{Number of channels in system with suppression of carriers}}{\text{Number of channels in system without suppression of carriers}}.$$

The numerical values of γ, calculated using the formulas presented above, are presented in Figures 8.10 and 8.11.

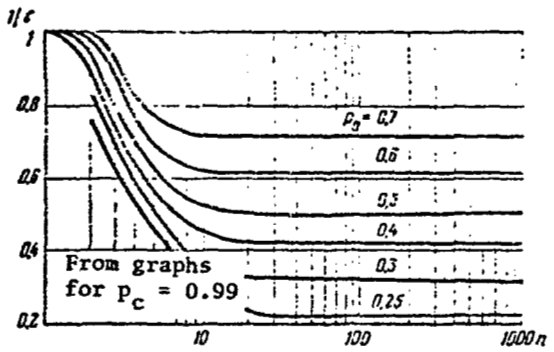


Figure 8.10. Estimation of Gain Provided by Statistical Multiplexing for p_c = 0.99

The figures indicate that in the example studied above where p_a = 0.5-0.3, suppression of carriers in a system consisting of $n \geq 10$ stations allows the throughput capacity to be increased by 3 to 4 times.

It should be noted that under actual conditions, some telephone channels may be occupied by transmission of secondary signals (telegraph, phototelegraph, etc.). Suppression of carriers in these channels, obviously, is impossible; therefore, the operational gain will be lower than the theoretically predicted gain.

Our analysis has assumed frequency multiplexing of FM signals. However, as we know, satellite communications systems may successfully use not only frequency modulation but also pulse code modulations (PCM). Many comparison of the two methods for lines (not systems) of satellite communication [8.8]-[8.10] have shown that with an equal band of frequencies and identical requirements for transmitted signal quality, approximately identical satellite transmitter powers are required in both cases.

Studies of the throughput capacity of a relay station in the multistation mode using PCM with frequency division multiplexing have not been published; nevertheless, the structure of PCM signals indicates that in the multistation mode this type of modulation would be preferable to FM, although the difference would not be great. /229

Going over to evaluation of the method of frequency division multiplexing as a whole, we must note its simplicity and flexibility. With frequency division multiplexing, the transition from communications lines (i.e., from a one-signal relay station operation mode) to a multistation system (i.e., to a multisignal mode) is relatively simple, requiring practically no modifidation of ground station apparatus. One important feature is the asynchronous mode of operation of the entire network, allowing combined operation of stations with antennas of different diameters and different exchange requirements.

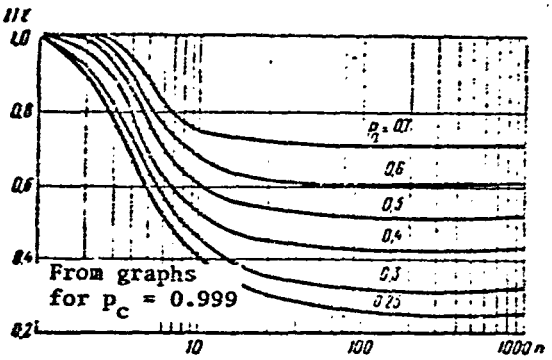


Figure 8.11. Estimation of Gain Achieved by Statistical Multiplexing for p_c = 0.999

The shortcoming of frequency division multiplexing noted above is the relatively low effectiveness of utilization of relay station power, leading to a reduction in its throughput capacity in the multistation mode. One additional complication is the necessity of regulating the power of ground stations (to avoid mutual suppression of signals) and the desirability of compensation for Doppler shifts in order to narrow the guard bands.

However, in spite of these shortcomings, the method of frequency multiplexing has been quite broadly used and at the present time is used in the USA as the main method both for internal satellite communications lines and for the international "Intelsat" system.

§ 8.4. Combination of Methods of SSB Modulation and FM with Frequency Division Multiplexing

/230

Let us study one version of frequency division multiplexing, based on the use of a combination of two modulation methods: SSB and PM (or FM). Single sideband modulation is used over the Earth-satellite section, where the power requirements are not so rigid, but flexibility in the redistribution of the number of telephone channels is required. Phase modulation is used in the sector between the satellite and Earth, where the power relationships are predominant.

The operation of a communications system, according to [8.11], is as follows (Figure 8.12). Each ground station radiates its signal in a strictly defined sector of the frequency spectrum (Figure 8.12, a, b, c). These subtrunk signals consist of sets of a certain number of telephone channels with frequency division multiplexing, for example standard 60 channel groups transmitted by the SSB method. Thus, the band of frequencies occupied by the signal of one station ΔF_{gr} is equal to the width of the frequency band of the group spectrum of the telephone channels and can be changed as the number of telephone channels allocated to the station changes.

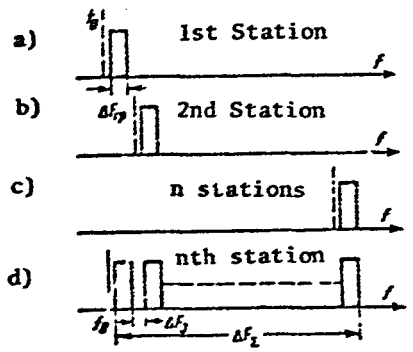


Figure 8.12. Frequency Division Multiplexing of SSB Signals

In the input circuits of the satellite receiver, the signals of the ground stations are added and form a super-group signal of the system (Figure 8.12d). In this case, if the frequency stability of the ground station is sufficient and there are no Doppler shifts (stationary satellite) or Doppler shifts are compensated during transmission, the guard bands between signals of stations with neighboring frequencies can be minimal and, furthermore, the total super-group spectrum will be near the standard spectrum for the corresponding number of telephone channels with frequency division multiplexing.

Subsequently, in the satellite relay station, the super-group SSB signal is converted to a wideband PM signal, for which the spectrum of modulating frequencies is equal to the width of the resultant super-group SSB spectrum (ΔF_{Σ} on Figure 8.12d). This signal is amplified and transmitted by the on-board transmitter. We note that since the converted PM signal is radiated at a single carrier frequency, the amplitude characteristic of the satellite transmitter can be essentially nonlinear; transient interference, /231 obviously, does not arise in this case.

Each of the ground stations receives the entire satellite signal, demodulates it to the super-group spectrum and then separates the telephone channels addressed to it.

In the simplest case, conversion of the SSB signal to a PM signal can be performed by adding it to a supplementary harmonic signal, the frequency of which corresponds to the frequency of the

suppressed SSB carrier. A simplified diagram of a relay station with this type of conversion is presented in Figure 8.13, a vector diagram explaining the conversion principle -- in Figure 8.14. The input SSB signals of the ground stations are fed through the antenna and the device matching reception and transmission to the input of the satellite receiver. The resulting signal is converted by a mixer, amplified by an IF amplifier and fed to the SSB-PM converter, where it is added with a reference voltage from oscillator O_2. The converted signal with PM is sent to a frequency multiplier, where the phase deviation is increased to the level necessary for interference-stable reception. Then the wideband PM signal is amplified by the final power amplifier, based on a traveling wave tube, and fed to the antenna through the matching device.

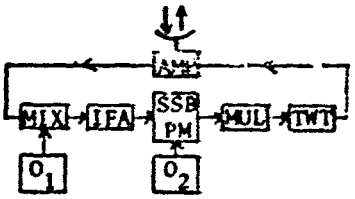


Figure 8.13. Simplified Block Diagram of Relay Station with SSB-PM Converter

In order to explain the quantitative relationships, let us study the conversion process. The input SSB signal with harmonic modulation of the voltage at frequency Ω can be recorded as

$$u_{in} = U_B \sin (\omega_0 + \Omega)t,$$

where ω_0 is the frequency of the suppressed carrier of the signal, while the supplementary voltage from oscillator O_2

$$u_0 = U_0 \sin \omega_0 t.$$

The sum of these voltages, the output voltage of the converter, can be represented as /232

$$u_{out} = U_0 \sqrt{1 + \beta^2 + 2\beta \cos \Omega t} \sin \left(\omega_0 t + \operatorname{arctg} \frac{\beta \sin \Omega t}{1 - \beta \cos \Omega t} \right). \quad (8.14)$$

Figure 8.14. Principle of Conversion of SSB to FM

where $\beta = U_B/U_0$ is the ratio of the amplitude of the input signal to that of the reference voltage.

It follows from (8.14) that the converted signal has amplitude and phase modulation. In our subsequent discussion, we will be interested only in the phase modulation, since it is this modulation which is the useful product of conversion. The amplitude modulation is later eliminated by a limiter.

The phase deviation of oscillation (8.14) is proportional to the amplitude of the input signal only in the first approximation, and the true rule of phase change of the output signal is nonharmonic, i.e., the process of conversion is accompanied by distortion of the shape of the modulating signal. In order to determine the harmonics, let us expand the expression for the phase of the output signal in a Fourier series:

$$\phi_{\text{out}}(t)=\sum_{n=1}^{\infty}\frac{\beta^n\sin n\Omega t}{n} \quad \text{where } \beta^2<1.$$

The deviation of phase of the output signal, determined by the first harmonic of the modulating signal, and the coefficients of the harmonics are respectively equal to:

$$\Delta\varphi_1=\beta,\ k_n=\frac{\beta^{n-1}}{n}=\frac{\Delta\varphi_1^{n-1}}{n}. \tag{8.15}$$

The formulas produced indicate that the converter in question actually converts the amplitude of the input signal to a phase deviation, but the possibilities for achievement of great deviations are limited by nonlinear distortions, which increase rapidly with increasing deviation.

In the case in question, when the band of frequencies at the input signal is essentially less than the carrier frequency, the primary role is played by the products of distortion of odd powers, since they are manifested a transient noise in the multichannel system. The presence of nonlinearity in the even powers has no influence on transient noise (this noise does not enter the band of the signal and can be filtered out).

In order to estimate practically possible values of phase deviation of a signal after the converter, let us determine the power of transient noise in a telephone channel of the multichannel system with frequency division multiplexing. According to the method of [8.5], the mean noise power in a telephone channel (weighted value) as applicable to the case in question can be determined as follows: /233

$$P_{nn} = 24 \frac{\Delta F_k k_c^2}{\Delta F_\Sigma} e^{-(2b_3 + 4 - 2b_m)} Y_3(\delta), \qquad (8.16)$$

where ΔF_k =3.1 KHz is the frequency band of the telephone channel

ΔF_Σ is the band of frequencies of the super-group spectrum of the SSB channel at the input of the converter;

k_c = 0.75 is the psophometric coefficient;

$b_3 = \ln(3/\Delta\phi_1^2)$ is a coefficient characterizing the attenuation of nonlinearity for the third harmonic, measured when a sine wave voltage is applied to the channel at a level equal to the peak level of the multichannel super-group signal;

ν is the peak factor of the super-group signal;

b_m is the mean level of the super-group signal;

$Y_3(\delta)$ is a coefficient considering the distribution of products of nonlinearity of third order in the linear spectrum; its value is presented in [8.5] in graphic form.

The results of calculations using (8.16), performed for the middle channel of a 600 channel system, are presented in Figure 8.15 as the dependence of noise power in the channel on peak phase deviation $\Delta\phi$ of the output signal; here also we find the dependence for the third harmonic coefficient calculated according to formula (8.15). /234

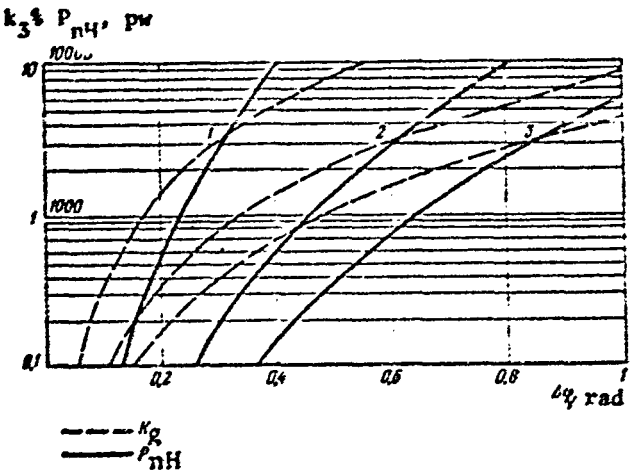


Figure 8.15. Power of Nonlinear Noise Upon Conversion of SSB to PM: 1, For Simple Converter; 2, For Differential Converter; 3, For Converter with Feedback

Acceptable values of transient noise power (on the order of 2-3 thousand pw) are produced with phase deviation not over 0.3 rad, whereas achievement of sufficiently high interference stability in reception requires full phase deviation on the order of 3-5 rad. Obviously, the necessary increase in phase deviation after conversion can be achieved only by frequency multiplication (see Figure 8.13), but the inclusion of multipliers with total multiplication factor 10-20 in the on-board equipment of the satellite complicates the relay station and reduces its reliability.

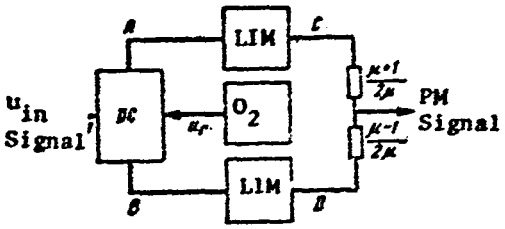


Figure 8.16. Block Diagram of Differential SSB-PM Converter

In connection with this, [8.11] suggested a more complex converter circuit, providing less nonlinearity of conversion. A diagram of this converter is shown in Figure 8.16, a vector diagram explaining its operation -- in Figure 8.17. The input SSB signal, following IF amplification, is sent to differential adder DC (actually made using a differential transformer); this device also receives supplementary oscillations from oscillator O_2. The frequency of the reference oscillation is the same as in the preceding converter, i.e., corresponds to the suppressed carrier of the super-group SSB signal.

The differential adder adds and subtracts the two oscillations, so that the output of the adder carries the sum-difference voltage. The voltage at point A is the sum $u_{in} + u_0$ and is determined by expression (8.14). The voltage at point B is the difference $u_{in} - u_0$; by analogy with (8.14), it can be represented in the same form. /235

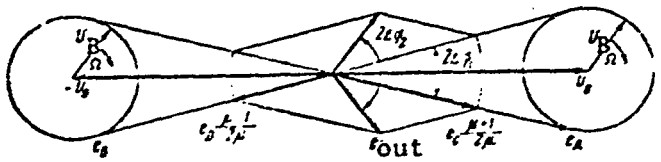


Figure 8.17. Vector Diagram of Differential SSB-PM Converter

Following the differential adder, both voltages, containing both amplitude and phase modulation, are sent to limiters, eliminating the parasitic amplitude modulation.

After limitation, the signals are added in the weight relationship $(\mu + 1)/(\mu - 1)$. This addition increases the deviation of the phase of the summary signal by μ times in comparison to the phase deviation of each of the signals added. Actually, by adding the two oscillations considering the weight factors, we can easily produce an expression for the phase of the output signal of the converter:

$$\phi_{out}(t) = \frac{1}{2} \arctan \frac{\beta^2 \sin 2\Omega t}{1 - \beta^2 \cos 2\Omega t} + \arctan$$

$$\arctan \mu \tan \frac{1}{2} \arctan \frac{2\beta}{1 - \beta^2} \sin \Omega t. \qquad (8.17)$$

The first component of this expression contains only even harmonics, the second component defines the useful phase deviation of the output signal and the odd harmonics. To estimate the linearity of the differential converter, let us expand (8.17) in a power series; after simplifications and ignoring powers over third power, we produce:

$$\Delta\varphi_1 \approx \mu\beta; \quad k_2 \approx \frac{\beta}{2\mu}; \quad k_3 \approx \frac{(\mu\beta)^2}{12}.$$

The corresponding dependences of the third harmonic factor and the power of transient noise in the telephone channel on phase deviation in the output signal are presented in Figure 8.15.

With the same value of phase deviation of the output signal, the differential converter has a four times lower nonlinear distortion factor for the third harmonic and, consequently, introduces 16 times less transient noise than the simple converter.

The primary shortcoming of both of these two converters is the nonlinearity of the conversion factor, limiting the possibility of producing the necessary phase deviation of the output signal and requiring its subsequent multiplication. As a means of increasing linearity of the conversion characteristic, [8.11] suggests introduction of phase feedback. The phase deviation and harmonic factors for $\beta^2 < 1$ in this case become:

$$\Delta\varphi_1 = \beta; \quad k_2 = 0; \quad k_3 = \frac{\beta^3}{24}. \qquad (8.18)$$

Comparing (8.18) with (8.15), we can see that the introduction of feedback of this type allows us to decrease nonlinear distortions of the simple converter to zero for the second harmonic and by a factor of 8 for the third. Furthermore, a converter with feedback is better than a differential converter, since with equal values of phase deviation of the output signal it introduces only half the nonlinear distortion. /236

Figure 8.18 shows a diagram of a converter realizing this relationship.

The relationships presented allow us to estimate the throughput capacity of a relay station with this combination of the two modulation methods. It is not difficult to see that with precise adjustment of frequencies and almost complete absence of protective guard bands, the throughput capacity of the relay station is independent of the number of ground stations. Actually, the structure and parameters of the multistation SSB signal at the input of the relay station are independent in principle of the number of ground stations sending the signals.

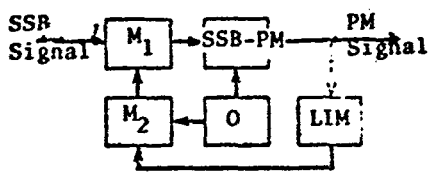


Figure 8.18. Block Diagram of SSB-PM Converter with Feedback

Thus, in this very interesting case, we produce the most effective system, in which multistation operation is not accompanied by a decrease in relay station capacity, i.e., Z(n) = const.

However, this system also has its shortcomings, preventing it from being used unconditionally and universally.

One of the most significant shortcomings is the necessity of extremely precise setting of frequencies of ground station transmitter. For example, with the current recommendations for instability of signal frequency in a telephone channel of 2 Hz, inaccuracy of adjustment of frequency for transmission in the 6 GHz band must be not over

$$\Delta = 2 \text{ Hz}/6 \cdot 10^9 \text{ Hz} \approx 3 \cdot 10^{-10},$$

which is extremely difficult to achieve.

One additional complication is that the lack of a gain in signal/noise ratio with SSB transmission means that the ground station transmitter power must be very high (up to 50 kw).

This method was experimentally tested using the ATS-1 satellite; the results produced indicated that it is effective for systems with stationary satellites.

§ 8.5. Division Multiplexing in Real Time

With time division multiplexing, all ground stations operate on the same frequency, but the transmitters of ground stations are switched on only for time τ at intervals of length T (Figure 8.19). When there is a rigid synchronization system, eliminating mutual superimposition of signals of neighboring stations, the multistation signal at the input of the relay station is a sequence of radio pulses with random filling frequencies. /237

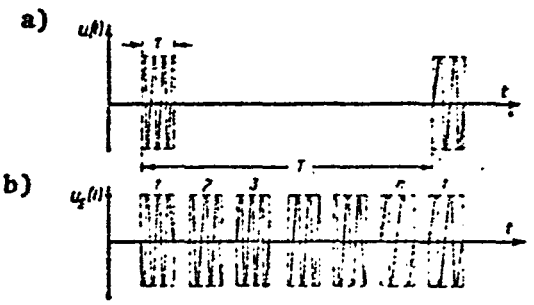


Figure 8.19. Principle of Time Division Multiplexing of Radio Signals

Since only one signal passes through the relay station at any one time with time division multiplexing, a significant portion of the effects characteristic for systems with frequency division multiplexing are absent (no losses of output power of the relay station, no mutual suppression of signals, no nonlinear transient interference). This indicates that the throughput capacity of a multistation system with time division multiplexing is independent of the number of stations.

However, this cannot be achieved in practice, since a new type of crosstalk interference appears, resulting from superposition of signals of neighboring stations due to inaccuracy of synchronization and the frequency-dependent properties of the relay station. Furthermore, the pulse operating mode results in the appearance of certain effects which also reduce the interference stability of reception and decrease the number of channels in the system. Finally, as the number of stations increases, overall time losses to guard time intervals also increase, reducing the throughput capacity of the system still further.

In order to explain this, let us study a simple time division multiplexing system operating in real time using FM signals.

The modulating signal at the transmitter of each ground station is fed continually and modulates the carrier frequency of the transmitter. The transmitters are connected in sequence for time τ with repetition frequency Ω_0, related to the highest frequency

of the modulating signal Ω_B by Kotel'nikov's theorem. Thus, the signal of each station (each subtrunk) is a sequence of radio pulses with filler frequencies changing according to the modulating signal (Figure 8.19a). The input of each receiver receives a sequence of pulses from various station (Figure 8.10b). The required signal is separated by time selection (for example, strobing) with subsequent frequency detection of the pulse sequence as shown on Figure 8.19a. /238

The inpulse operation of the transmitters causes the spectrum of the FM signal to expand and, consequently, occupy a broader band of frequencies.

The analytic expression for an impulse FM signal with sine-wave frequency modulation is

$$U(t)=U_0\cos[\omega_0 t+m\sin(\Omega t+\varphi)]\sum_{l=-\infty}^{\infty} f(t-t_l),$$

where m is the frequency modulation index;
f(t) is a function describing an individual pulse;
$t_1 = 1T - t_0$ is the time from the beginning of the time reading t_0.

As was shown in [8.12], the spectrum of this signal is determined by the formula

$$S(\omega)=\frac{U_0}{2T}\sum_{k=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\Phi(k\Omega_0)J_n(m)\cos[(\omega_0+n\Omega+k\Omega_0)t+n\varphi-\alpha_0], \qquad (8.19)$$

where $\Phi(k\Omega_0)$ is the modulus of the spectrum of the pulse sequence;
$J_n(m)$ is the Bessel function of the real argument.

The location of the frequency components is shown on Figure 8.20, which shows the spectra of a continuous FM signal, a sequence of rectangular RF pulses with unchanged filler frequency and the spectrum of a pulse FM signal with time division multiplexing.

Figure 8.20c indicates that in the spectrum of the pulse FM signal around each discrete component of the spectrum in the pulse sequence are components of the spectrum of the FM signal.

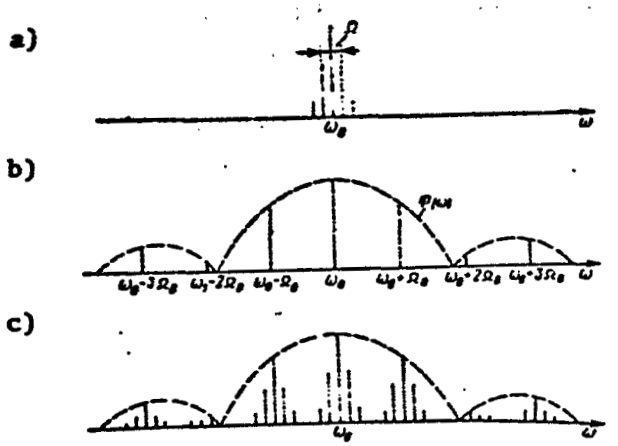


Figure 8.20. Spectrum of Pulse FM Signal

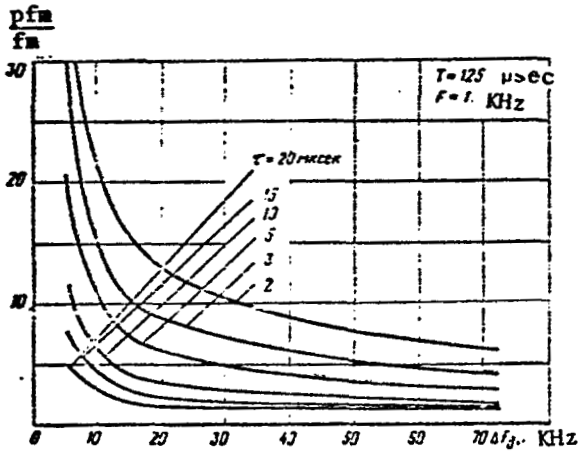


Figure 8.21. Expansion of Spectrum of Pulse FM Signal in Comparison to Spectrum of Continuous FM Signal

Consequently, the spectrum of the pulse FM signal is broader than the spectrum of a continuous FM signal by a quantity determined by the width of the spectrum of the switching pulse sequence. With a large number of stations, this width may be significant and may require a significant increase in the frequency band of the relay station and the ground station receivers.

Figure 8.21 shows experimental curves characterizing the expansion of the spectrum of an FM signal for pulse operation in comparison with the spectrum of a continuous FM signal.

Evaluation of the power capabilities of this method [8.13] shows that the potential interference stability is near the potential interference stability of continuous FM.

Actually, according to the method described in [8.1], the spectral noise density at the output of an ideal receiver for the pulse FM signal /240

$$\sigma^2 = \frac{\sigma_0^2 (2\pi f)^2}{\left[\frac{dU(t)}{d\varphi}\right]^2} = \frac{2\sigma_0^2 f^2}{U_д^2 \Delta f_д^2} \frac{T}{\tau},$$

where σ_0 is the spectral noise density at the input of the receiver.

Integration of σ^2 in the band of frequencies of the LF filter gives us

$$P_n = \int_0^{F_B} \sigma^2 df = \frac{2}{3} \frac{\sigma_0^2 F_B}{U_0^2 m^2} \frac{T}{\tau},$$

where $m = \Delta f_d / F_B$ is the frequency modulation index.

The signal/noise ratio at the output of the receiver

$$(P_s/P_n)_{out} = \frac{3}{4} \frac{U_0^2 m^2}{\sigma_0^2 F_B} \frac{\tau}{T},$$

while the ratio of mean powers of signal and noise at the input of the receiver

$$(P_s/P_b)_{in} = \frac{U_0^2}{2\sigma_0^2 \Delta f} \frac{\tau}{T}.$$

Consequently, the gain of FM, characterizing the interference stability of an ideal receiver,

$$B_{id} = \frac{(P_s/P_n)_{out}}{(P_s/P_n)_{in}} = 3m^2(\Delta f/2F_B) = B_{id}(FM),$$

i.e., it coincides with the gain for continuous FM.

Correspondingly, comparing the two methods with equal peak (not mean) signal powers at the input of the receiver, we produce

$$B_{id}^* = 3m^2(\Delta f/2F_B)(\tau/T) = B_{id}(FM)\frac{1}{\beta_{id}}. \qquad (8.20)$$

Thus, in the pulse mode the signal/noise ratio at the output of the ideal receiver should be τ/T times less than for a continuous receiver.

In [8.13], the actual interference stability was determined for an FM receiver with a standard FM demodulator and a time selector set at the input (or output) of the demodulator.

As was shown in [8.13], the noise spectrum at the output of the time selector, produced by scanning the noise spectrum of the output of the frequency detector and the spectrum of the sequence of selecting pulses /241

$$\varpi(f)=\left(\frac{\tau}{T}\right)^2\frac{A(f)^2+A(-f)^2}{2}+$$
$$+\frac{2\tau^2}{T^2}\sum_{l}^{\infty}\left\{\frac{\sin k\pi\frac{\tau}{T}}{k\pi\frac{\tau}{T}}\right\}^2\frac{A(kf_0-f)^2+A(kf_0+f)^2}{2}, \tag{8.21}$$

while the noise, in the band of the LF filter (in a telephone channel)

$$P_{nk}=2\frac{\tau^2}{T^2}A\frac{F_n^3}{3}\sum_{0}^{l}\left[\frac{\sin k\pi\frac{\tau}{T}}{k\pi\frac{\tau}{T}}\right]^2(12k^2+1), \tag{8.22}$$

where $l = \Delta f/2f_0$ is the ratio of the band of frequencies of the receiver to twice the cycle frequency.

Accordingly, the gain in the actual receiver

$$B_p^* = B_{id}(FM)\frac{1}{\beta_p},$$

where

$$\beta_p=2\sum_{0}^{l}\left[\frac{\sin k\pi\frac{\tau}{T}}{k\pi\frac{\tau}{T}}\right]^2(12k^2+1).$$

The quantitative characteristic β_p is presented on Figure 8.23, which shows the dependence of β_p on the number of ground stations with various widths of the frequency band of a ground station receiver, characterized by coefficient $\mu = \Delta f\tau/2$; here also we see the dependence $\beta_{id} = f(n)$, calculated from formula (8.20).

It follows from Figure 8.22 that even with a small number of stations, the interference stability of an actual receiver is 20-30 db worse than the ideal level. This can be explained physically by Figure 8.23, showing the noise spectrum at the output of an actual pulse FM signal receiver, calculated using formula (8.21). When the parabolic noise spectrum at the output of the frequency detector is combined with the spectrum of the sequence of selecting pulses, a pair of side parabolic noise spectra is formed around each component of the spectrum of the pulses (Figure 8.23, solid line shows side spectra around zeroth component of pulse spectrum; the dotted and dot-dash lines show the side spectra around the first and second components). In this case, the band of the low frequency filter includes not only the main component of parabolic noise, but also a portion of the lower side components from some of the first discrete components. Since the width of each side component is determined by the band width of the receiver, the number of discrete components ℓ contributing to the noise in the band of the LF filter is proportional to the band width of the receiver, while the total power of all noise components is determined by formula (8.22). /242

Thus, an ordinary FM signal receiver with the addition of a time selector does not realize the potential capabilities of pulse FM in the multistation mode.

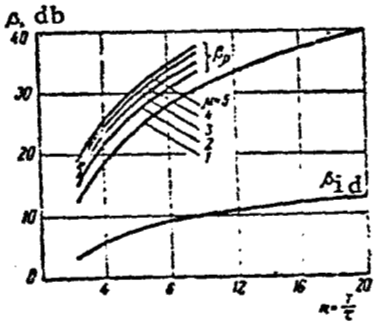


Figure 8.22. Quantitative Estimate of Loss in Interference Stability with Multistation Access with Time Division Multiplexing in Real Time

The simplest method which can be used to increase the interference stability of an actual FM receiver of this type consists of increasing the cycle frequency to $f_0 \geq 2\Delta f_d = 2mF_B$. In this case, the distance between discrete components of the pulse spectrum will be broader than the spectrum of the continuous FM signal (Figure 8.24), and the spectral components will not overlap. This fact allows a filter with

bandwidth $\Delta f = 2\Delta f_d$ to be installed in the receiver, to separate the basic component of the pulse FM spectrum. It can be shown that the interference stability of the receiver in this case is near the potential stability, but the increase in the cycle frequency results in a significant expansion of the signal spectrum which, in the final analysis, has an unfavorable influence on the throughput capacity of the relay station. /243

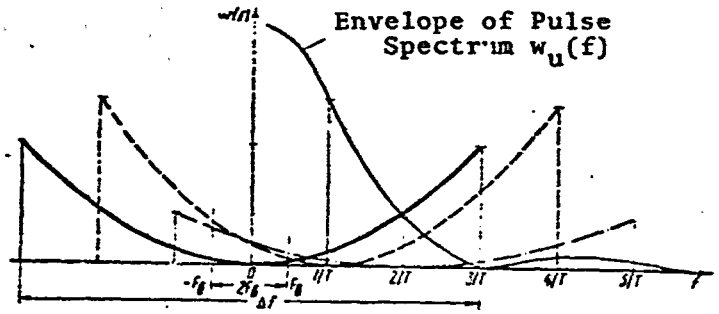


Figure 8.23. Noise Spectrum at Output of Actual FM Pulse Signal Receiver

More radical measures for elimination of the loss in interference stability require the use of special demodulators, for example a demodulator which separates the band of frequencies [8.14] or a demodulator with a tracking oscillator [8.15]; in the latter case, the potential interference stability of the modulation method is almost fully realized, avoiding losses related to expansion of the spectrum.

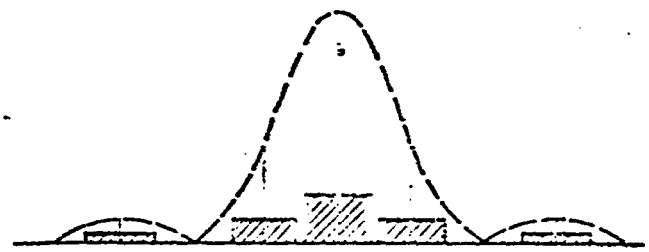


Figure 8.24. Spectrum of Pulse FM Signal with Increased Cycle Frequency

Evaluating this version as a whole, we must note that with a comparatively low number of stations, apparently, extremely good

results can be achieved. As the number of stations rises, the problems of time synchronization become quite significant and may limit the applicability of this method significantly.

§ 8.6. Time Division Multiplexing with Transformation of Time Scale

The method of time division of channels with transformation of time was suggested in [8.16] and described in [8.17]. The essence of this method as applicable to a multistation satellite communications system can be briefly explained using the time diagram shown in Figure 8.25.

The input low frequency signal, either a single-channel telephone message or a group of several telephone messages (formed using standard group multiplexing equipment used in radio relay lines) for each ground station is subjected to time compression, for example using a tape recorder with delayed recording and accelerated reproduction. These time-compressed signals (shown by the dotted line on the figure) are sent to the FM transmitter and radiated during time τ with period T[1]. Rigid synchronization of the operation of all ground stations and introduction of protective intervals τ_g assures that there will be no mutual interference between neighboring signals. Thus, the multistation signal passing through the relay station is a sequence of radio frequency pulses with filling frequency changing over interval τ according to the envelope of the messages transmitted. /244

The reception and separation of signals at the ground stations is performed by methods of frequency detection and time selection with subsequent time expansion. In order to increase the interference stability, the transmission and reception channels include linear predistorting and restoring circuits.

By analogy with the investigation of frequency division multiplexing, [8.18] calculates the basic parameters of the system in question, consisting of n ground stations, each with throughput capacity N' telephone channels, establishing the dependence $N' = f(n)$.

As before, the initial condition is constancy of noise power in the telephone channel

$$P_{nk} = P_{nt} + P_{nn} + P_{nc} = \text{const},$$

[1] In contrast to the time system described earlier, the duration of period T is not related to the highest frequency of the messages transmitted and is limited only by the permissible time delay in transmission of information.

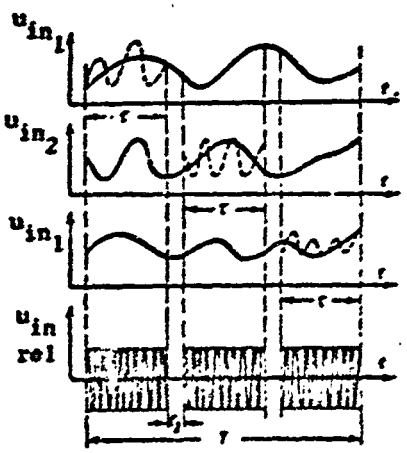


Figure 8.25. Time Division Multiplexing with Time Transformation

where P_{nt} is the power of thermal noise;

P_{nn} is the power of nonlinear noise;

P_{nc} is the power of transient noise.

The power of thermal noise in a telephone channel can be determined from a known formula [8.5] for the case when time compression at each ground station is performed for a standard group of N telephone channels, considering transformation of the telephone channel spectrum

$$P_{nt} = 10^9 \frac{kT_{\Sigma_g} \Delta F_k k_c^2 L_\Sigma}{P_\sigma \Delta_{ct}^\beta \Delta f_d^2} \gamma^3 F_k^2 e^{2b_0}, \text{ pw} \quad (8.23)$$

where Δ_{ct} is the gain of thermal noise resulting from introduction of linear predistortion; /245

b_0 is the difference between the maximum level of a multichannel message and the measurement level of one channel, nep;

γ is the time compression factor;

β is a coefficient, considering the increase in noise in a telephone channel due to time selection in the receiver.

Let us study the quantities included in (8.23). Some of them, such as P_s, T_{Σ_g}, L_Σ, etc. are independent of the number of ground stations and determined by the initial parameters of the system; others are functions of n or N'.

Thus, the time compression factor

$$\gamma = \frac{T}{\frac{T}{n} - \tau_g} = nq_t, \quad (8.24)$$

where $q_t = (\tau + \tau_g)/\tau$ is the duty factor of radio pulses in the multistation signal.

The frequency of the upper channel

$$f_k = N'F_B a = N'a3\ 4\ \text{KHz}, \tag{8.25}$$

where a = 1.25-1.4 is a coefficient considering the presence of the guard intervals between neighboring channels with group frequency division multiplexing.

Calculation of coefficient β can be based on results presented in the preceding case (using them, we can show that where $T \gg 1/2F_B$, the value of $\beta \to 1$).

Finally, let us assume $k_c = 0.75$ (for a rectangular noise spectrum) and $\Delta_{ct} = 4$ db[1] -- the value characteristic for radio relay lines.

Substituting all these values into (8.23) for a = 1.3 and β = 1, we produce

$$P_{nt} = 2 \cdot 10^{-4} \frac{T_\Sigma L_\Sigma}{P_\sigma \Delta f_d^2} n^3 q_t^3 (N')^2 e^{2b_0}, \text{ pw} \tag{8.26}$$

The noise power of nonlinear transitions arising as the group FM signal passes through a line with nonlinear phase characteristic is also determined by transformation of the spectrum of the telephone channels by a known formula [8.5]: /246

$$P_{nn} = 10^9 \frac{\Delta F_k k_c^2}{\Delta F_{gr}} 4\pi^2 n^2 q_t^2 f_k^2 \left[\frac{(\delta t)^2}{4\Delta_{2H}} e^{4b_m - 2b_0} y_2(\sigma) + \right.$$

$$\left. + \frac{(\Delta\tau)^2}{6\Delta_{3H}} e^{6b_m - 4b_0} y_3(\sigma)\right], \text{ pw} \tag{8.27}$$

where ΔF_{gr} is the group signal frequency band;

$\delta\tau$, $\Delta\tau$ are the asymmetry and unevenness of the group delay time characteristics, measured with frequency deviation equal to the peak deviation of the FM signal;

$y_2(\sigma)$, $y_3(\sigma)$ are the distribution functions of the products of second and third order nonlinearity in the group spectrum [8.5];

[1] With frequency characteristic of predistortions as recommended by ICCR.

Δ_{2H}, Δ_{3H} is the gain with respect to second and third order nonlinear noises, resulting from the introduction of predistortion;
b_m is the mean level of the multichannel message.

Substituting (8.25) into (8.27) and assuming that for the upper telephone channel $\Delta_{2H} = 3.5^1$, $\Delta_{3H} = 2.3^1$, $y_2(\sigma) = y_3(\sigma) = 0.5$, we produce

$$P_{nn} = 1{,}1 \cdot 10^{16} n^2 q_i^2 N' \left[(\delta\tau)^2 e^{4b_m - 2b_0} + (\Delta\tau)^2 e^{6b_m - 4b_0}\right], \text{ pW}. \quad (8.28)$$

The thermal and nonlinear noise calculated by formulas (8.27) and (8.28) make up the main portion of the noise in the telephone channels of this system. Noise resulting from mutual superimposition of ground station signals can be reduced as much as desired by selecting the guard interval τ_g. Other noise components, such as noise from competing services, etc., have no influence on the results of investigation in the first approximation and therefore will not be considered.

Thus, general formula (8.23), relating all of the power parameters of the system in question, becomes

$$P_{nk} \approx An^3 q_i^3 (N')^2 e^{2b_0} + Bn^3 q_i^2 N' \left[e^{4b_{cp} - 2b_0} + \left(\frac{\Delta\tau}{\delta\tau}\right)^2 e^{6b_{cp} - 4b_0}\right]. \quad (8.29)$$

Formula (8.29) is correct if the group signal of each of the ground stations consists of at least three telephone channels. If each station radiates single-channel signals, this formula is inapplicable.

Actually, when single-channel messages are transmitted, nonlinear noise is absent in principle and the nonlinearity of the phase characteristic, as we know, leads to appearance of nonlinear distortions in the telephone channel.

As was shown in [8.5], the coefficients of nonlinear distortions for the second and third harmonics /247

$$k_2 = \delta\tau\pi F, \quad k_3 = \frac{\pi F}{2}\Delta\tau.$$

[1] With the predistortion frequency characteristic recommended by ICCR.

where F is the frequency of the measurement tone in the telephone channel (usually 1000 Hz).

In the system in question with time compression $F' = \gamma F$, and the expression for the harmonic factor becomes

$$k_f = 3{,}14 \cdot 10^3 n q_t \sqrt{(\delta\tau)^2 + \left(\frac{\Delta\tau}{2}\right)^2}.$$

The required nonlinear distortion factor (residual attenuation should be constant with accuracy ±0.035 nep) in a telephone channel must be provided with an overload level of 0.4 nep. Thus, when single-channel messages are transmitted, the peak deviation of frequency corresponding to the measured level in channel Δf_{d0} should be less than the peak deviation with transmission of multichannel messages, i.e.

$$\Delta f_{d0} = \frac{\Delta f_d}{p}, \text{ where } p \geq 1.5.$$

The expression for thermal noise upon transmission of single-channel messages, by analogy with (8.20), can be written as

$$P'_{nt} = \frac{2}{3} 10^9 \frac{kT_{\Sigma g} L_{\Sigma g} k_c^2 F_B^3 n^3 q_t^3}{P_\sigma \beta \Delta f_{d0}^2}$$

or, if $k_c = 0.65$ for the triangular noise spectrum in the channel,

$$P'_{nt} = 1.5 \cdot 10^{-4} \frac{T_{\Sigma g} L_\Sigma}{P_\sigma \beta \Delta f_d^2} n^3 q_t^3 p^2.$$

The set of formulas produced allows us to perform power calculation of systems with time division mutliplexing and determine the dependence of the throughput capacity of the system on the number of ground stations. Let us calculate the function $N' = f(n)$ if the parameters of the relay station and ground stations remain unchanged.

The results of calculations for several values of q_t are presented in Figure 8.26. They indicate that the throughput capacity of the system in question decreases slightly with increasing

numbers of ground stations. For example, a relay station capable of continuously transmitting 600 telephone channels between two stations, where $n = 23$ and $q_t = 1$ can transmit 276 telephone channels (12 channels from each station). When single-channel telephone messages are multiplexed, the throughput capacity of the system is minimal and is about 200 channels, i.e, approximately 30% of the initial level. /248

It should be noted that of the two systems in question for multistation access with time division multiplexing, the latter (with time transformation) is more effective and provides more complete utilization of the relay station capacity.

Comparison with Figures 8.26 and 8.27shows that with equal conditions, systems with time division multiplexing of FM signals are more effective than systems with frequency division mutliplexing of FM signals, for which the dependence of throughput capacity of the relay station on number of ground stations is more sharply expressed.

We must add that, since the repetition interval in a system with time transformation is not related so closely to the highest modulating frequency as in a system operating in real time, problems of synchronization can be significantly facilitated by selecting rather large values of T.

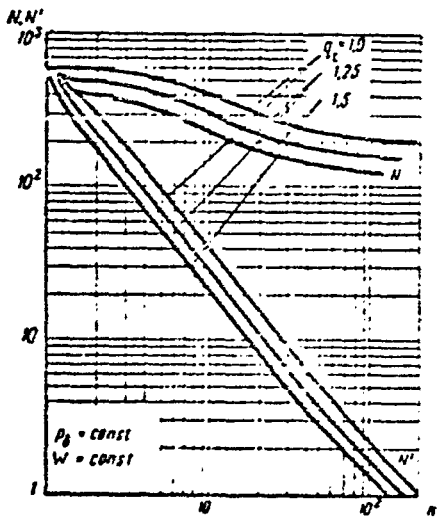


Figure 8.26. Throughput Capacity of a Relay Station with Time Division Multiplexing of FM Signals

Going over to an evaluation of the method of time division multiplexing as a whole, we must note its high effectiveness, resulting from its complete utilization of the power of the relay station and the absence of nonlinear effects, so characteristic for frequency division multiplexing.

However, the necessity of rigid time synchronization complicates its practical realization, particularly with a large number of ground stations. Some difficulties are also caused by joint operation of stations with different energy potentials and different throughput capacities. This problem can be solved in principle by uneven time compression

(signals of small ground stations compressed in time to a lesser extent), but this represents certain design and organizational difficulties.

Nevertheless, the high effectiveness of time division multiplexing is a decisive factor which will apparently result in broad utilization of this method in the future.

§ 8.7. Orthogonal Waveshape Multiplexing

/249

The construction of communications system in which the signals of ground stations are radiated simultaneously and occupy the same frequency spectrum is possible due to selection of a set of orthogonal signals which differ in waveshape.

The good mutual correlation properties of such signals allow them to be separated at the reception point using matching filters or correlators. However, the presence of interfering stations and the nonlinearity of the characteristics of the relay station unavoidably reduce the throughput capacity of the system during multistation operation. Multistation communications systems with complex (noise-like) signals were studied in [8.19] and [8.20].

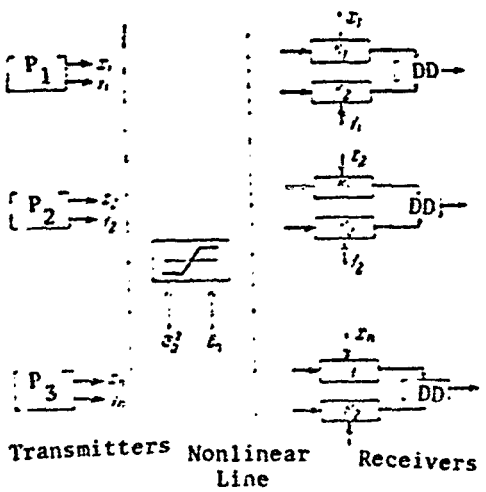


Figure 8.27. Block Diagram of Multistation System with Waveshape Multiplexing of Signals

A block diagram of one possible system using such signals is presented in Figure 8.27. Each of the end stations transmits information in binary form. A 1 signal corresponds to the complex signal $x_n = S_n^1(t)$, a 0 signal corresponds to the signal $y_n = S_n^0(t)$. White Gaussian noise N(t) with spectral density σ_0^2 and interference from neighboring stations $b_{n-1}(t)$, resulting from the incomplete orthogonality of the signals in the system, are also present in the channel. The energy E (and duration T) of the signals of all stations are identical, equal to

/250

$$E=\frac{1}{2}\int_0^T |S_n(t)|^2\,dt=\text{const.}$$

The mutual correlation factor of the signals received

$$\frac{1}{\tau}\int_0^T S_i^k(t)\,S_i^m(t-\tau)\,dt=\begin{cases}Er_i e^{i\alpha_i} & \text{where } k=m,\ l\neq i,\\ 0 & \text{where } k\neq m \text{ for all } l \text{ and } i,\end{cases}$$

where k and m are indices, taking on values of 0 or 1, depending on the nature of the signal transmitted by the ground station.

Each receiver has two filters K_1 and K_2, matched to transmitting signals $S_n^1(t)$ and $S_0^n(t)$ respectively. The voltage at the output of each filter

$$U_1=\frac{1}{2}\int_0^T S(t)\,S_n^{1*}(t)\,dt.$$

$$U_2=\frac{1}{2}\int_0^T S(t)\,S_n^{0*}(t)\,dt.$$

where S(t) is the resultant signal reaching the input of the receivers:

$$S(t)=S_n^m(t)\,e^{i\varphi_n}+N(t)+\sum_{i=1}^{n-1} S_i^k(t)\,e^{i\varphi_i}.$$

Decision device DD utilizes the sign of the difference $U_1^2 - U_2^2$ to determine the signal transmitted and correspondingly decides on either a 1 or a 0.

As was shown in [8.20], the full probability of error in reception is related to the parameters of the signals in the system by the following formula

$$P_{er}=\frac{1}{2}\exp\left(\frac{h_0^2}{2+r^2h_0^2n}\right), \qquad (8.30)$$

where $h_0^2 = \frac{E}{\sigma_0^2}$ is the ratio of the energy of one signal to the spectral density of the noise;

r is the maximum value of the modulus of the mutual correlation factor between signals;

n is the number of stations in the system.

Figures 8.28 and 8.29 show curves calculated by formula (8.30); they allow practical design of the system and determination of the throughput capacity for multistation operation.

The influence of nonlinearity of the amplitude characteristic of the relay station on interference stability and throughput capacity of the system was studied in [8.21]. The relay station model used was a limiter with non-zero threshold and amplitude characteristic as shown in Figure 8.27: /251

$$f(X) = \begin{cases} U_0 & \text{where } X \geqslant U_0. \\ X & \text{where } |X| < U_0. \\ -U_0 & \text{where } X \leqslant -U_0. \end{cases}$$

The models of signals studied were sectors of Gaussian noise of duration T, occupying total frequency band W.

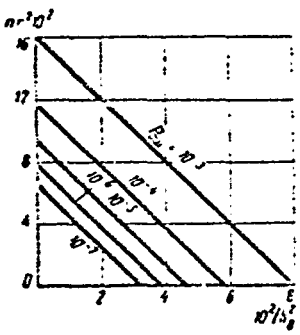


Figure 8.28. Correlation Factor of Signals Required for Fixed Probability of Error in Reception

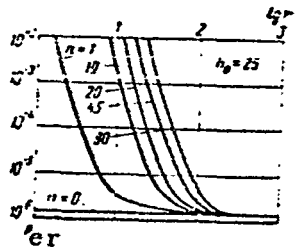


Figure 8.29. Probability of Error in Reception as a Function of Correlation Factor of Signals with Various Numbers of Ground Stations

The influence of amplitude nonlinearity is evaluated on the basis of the value and nature of changes in detection parameter

$$d = \frac{M(U_1^2 - U_2^2)}{\sqrt{D(U_1^2 - U_2^2)}}$$

where $M(U_1^2 - U_2^2)$ is the mathematical expectation of the difference of output voltages of matched filters K_1 and K_2;

$D(U_1^2 - U_2^2)$ is the dispersion of the same difference.

As we know, the detection parameter is related to the probability of error in reception by the formula

$$P_{er} = \frac{1}{2\pi}\int_d^\infty e^{-\frac{z^2}{2}}dz.$$

This dependence is illustrated by Figure 8.30.

As was shown in [8.21], when operating in the linear mode, when the summary level of signals of all stations and noise is less than the bend point on the amplitude characteristic of the relay station, detection parameter d_1 is /252

$$d_1 = \sqrt{\frac{WT}{n + \frac{WT}{h_0^2} + \frac{1}{2}}}.$$

Figure 8.31 shows the dependence of the detection parameter in a linear system with number of stations n = 10, 25 and 100 on base of signals used WT where $h_0^2 = 20$.

In the nonlinear mode, the multistation signal is limited, accompanied by distortion of signal shape and appearance of crosstalk interference. This increases the probability of reception error and reduces the throughput capacity of the relay station.

Figure 8.32 shows the ratio of detection parameters in linear d_1 and nonlinear d_0 modes as functions of percentage limitation, characterized by the ratio of limiter threshold U_0 to the mean square value of the sum of signals of all ground stations plus noise σ_Σ.

As follows from Figure 8.32, even with high percentage limitation ($U_0/\sigma_\Sigma = 0.2$), the value of the detection parameter decreases by not over 0.8 db, which indicates the stability of complex signals

to amplitude limitation. This stability is explained not only by the structure of noise-like signals, but also by the effective method of processing used, assuming that models of the signals transmitted are available at the reception point.

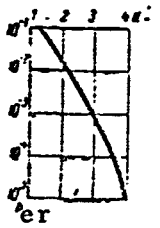


Figure 8.30. Reception Error as a Function of Detection Parameter

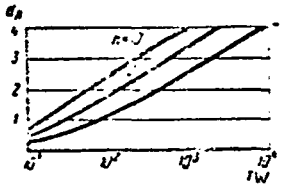


Figure 8.31. Detection Parameter as a Function of Base of Signals Used in Linear Multistation System

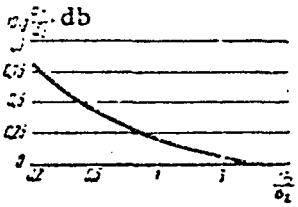


Figure 8.32. Estimating the Reduction in Detection Parameter with Amplitude Limitation of Multistation Signal

§ 8.8. Information Evaluation of Multiplexing Methods

/253-

The results produced above are related to the selection of specific signals; therefore, they provide no basis for prediction of the throughput capacity of the channel with other signals and other modulation methods.

Due to this, there is doubtless interest in general theoretical limiting estimates based on criteria from information theory, allowing us to determine the throughput capacity of a communications channel with optimal signals (with optimal coding) and study the nature of its changes for arbitrarily small reception error probabilities.

Knowledge of these estimates allows us to compare the methods of multiplexing of a channel under optimal conditions, to determine the most effective method and establish the limit of utilization of the throughput capacity of the channel with multistation operation. Comparison of results of the information evaluation with results produced earlier for practical signals allows us to determine how far the signals used are from the optimal and the loss in throughput capcity of the channel which results from their use.

As we know, the throughput capacity of a channel refers to the maximum rate of information transmission:

$$C = R_{max} = \lim_{T \to \infty} 2W\,[H_{max}(y) - H(n)], \qquad (8.31)$$

where H(y) is the entropy of one element of the signal received;
H(n) is the entropy of an element of noise;
W is the band of frequencies of the communications channel.

It follows from (8.31) that the throughput capacity is determined not only by the parameters of the channel (band of frequencies W), but also by the parameters of the signal and noise interacting with it, represented in (8.31) by the entropies H(y) and H(n). Since with fixed nature and power of noise, its entropy is independent of the signal parameters, maximization of R is possible only by selection of a set of signals providing the maximum of entropy H(y). If in this case the parameters of the communications channel place certain limitations on the characteristics of the signals, entropy H(y) should be maximized considering these limitations.

The greatest practical significance is that of determining the throughput capacity of the channels with limited mean power and with limited peak power of signals.

As was shown in [8.22], when there is a limitation of mean power, the highest entropy is that of the normal distribution and the transmission rate is at its maximum with signals similar to white noise; the throughput capacity of this channel

$$C = W \log\left(1 + \frac{P_s}{P_n}\right). \qquad (8.32)$$

Formula (8.32) determines the greatest throughput capacity which can be achieved upon transmission of signals with limited mean power with any, arbitrarily small, error probability. /254

For channels with limited peak power, we cannot produce a solution as simple as that for channels with limited mean power. Work [8.22] gives the boundaries for estimation of throughput capacity. Thus, for all values of S/P_n (where S is the peak power of the signal), the lower boundary of the throughput capacity of the channel is determined by the inequality

$$C \geqslant W \log \frac{2}{\pi e^2} \frac{S}{P_n}; \qquad (8.33)$$

the upper boundary for sufficiently high S/P_n

$$C < W \log\left(1 + \frac{2}{\pi e}\frac{S}{P_n}\right)(1+\varepsilon), \tag{8.34}$$

where $\varepsilon \to 0$ where $S/P_n \to \infty$.

In order to achieve the upper boundary of throughput capacity, one must select a set of signals with evenly distributed probabilities, since when the peak values are limited, the maximum entropy is that of a signal with a distribution which is even over the interval from $-\sqrt{S}$ to $+\sqrt{S}$. The peak factor of this distribution is equal to $\sqrt{3}$; therefore, the mean power $P_c = S/3$.

Thus, for two channels with different limitations placed on the power of the signals transmitted, essentially different expressions are produced, determining the throughput capacity in the presence of white noise.

It follows from the formulas which we have presented that with all signal/noise ratios, the greatest throughput capacity is that of a channel with limited mean power. It is approached by the upper boundary of throughput capacity of a channel with limited peak power; the lower boundary lies in the area of significantly lower signal/noise ratios.

In order to use the formulas presented to estimate the throughput capacity of channels with various methods of multiplexing, we must determine the type of limitation on power existing in the channel in question.

It is difficult to answer this question unambiguously, since the nature of the limitation depends on the type of signals used and the conditions of utilization of the channel. Strictly speaking, in a satellite communications channel we should consider the limited peak power, since the system has a relay station with an amplitude characteristic close to the characteristics of a rigid limiter, and the difficult power conditions characteristic /255
for space communications lines force us to use the relay station power as completely as possible. However, in many cases (for example in the transmission of noise-like signals), the possibility appears of reducing the output power of the relay station; furthermore, the structure of noise-like signals is such that the presence of amplitude limitation does not significantly increase the number of reception errors, allowing operation in the nonlinear area of the amplitude characteristic of the relay station. The limitation on peak power of the signal in this case can be ignored and the throughput capacity can be approximately determined using the formula produced for a channel with limited mean power.

Thus, since the type of power limitation depends on the specific conditions, it is expedient to study both limiting modes of operation of the channel, estimating the degree of practical approach to either boundary in each individual case.

Let us determine the throughput capacity of a channel with frequency division multiplexing.

Let us study a communications channel with bands of frequencies W, in which there is a certain signal with mean power P_s and peak power S; the channel also contains additive white noise with spectral density σ_0^2.

With frequency division multiplexing, this overall channel is divided into n partial channels (subtrunks) with frequency band width W_i, mean power P_i and peak power S_i. In order to eliminate mutual interference, guard bands W_g are introduced to separate the signals. Thus

$$W=\sum_{i=1}^{n}(W_i+W_g),\quad \frac{W}{W_i}=q_W,\quad W_i=q_W\sum_{i=1}^{n}W_i, \tag{8.35}$$

where $q_W = (W_i + W_g)/W_i$ is the factor characterizing the utilization of the frequency band of the channel.

If all partial channels are identical

$$W=q_W nW_i. \tag{8.36}$$

The problem consists in determination of the throughput capacity of the sum of partial channels and its comparison with the throughput capacity of the initial channel without multiplexing.

In a channel with limited mean power

$$P_S=\sum_{i=1}^{n}P_i, \tag{8.37}$$

and with identical partial channels $P_i = P_s/n$. The noise power in a partial channel is equal to $\sigma_0^2 W_i$. Consequently, the throughput capacity of a partial channel, according to (8.32), will be /256

$$C_i = W_i \log\left(1 + \frac{P_i}{\sigma_0^2 W_i}\right),$$

while the throughput capacity of the sum of independent partial channels

$$C_\Sigma = \sum_{i=1}^{n} W_i \log\left(1 + \frac{P_i}{\sigma_0^2 W_i}\right).$$

Let us express C_Σ through the parameters of a multistation signal. Since the greatest throughput capacity C_Σ is achieved with identical partial channels, considering (8.36) and (8.37) we produce

$$C_\Sigma = \frac{W}{q_W} \log\left(1 + \frac{P_s q_W}{\sigma_0^2 W}\right). \tag{8.38}$$

Comparing (8.38) and (8.32), it is not difficult to note that with limited mean signal power, the throughput capacity is independent of the number of channels multiplexed and losses in throughput capacity with multistation operation are determined only by the width of the guard frequency bands, characterized by factor q_W. This conclusion is correct if q_W is independent of the number of channels. However, in the general case, there may be such a dependence and it must be considered, although it is weakly expressed.

This situation is quite different in a channel with limitation on peak power. In this case, the peak power is divided among the channels and with identical partial channels

$$S_i = \frac{S v_i^2}{\sum v_i^2}. \tag{8.39}$$

where ν_i is the peak factor of the signal in a partial channel;

ν is the peak factor of the multistation signal.

With sufficiently high signal/noise ratios, the throughput capacity of the sum of independent partial channels, considering (8.39), is equal to

$$\frac{W}{q_W}\log\left(\frac{2}{\pi e^2}\frac{Sq_W\nu_i^2}{\nu^2\sigma_0^2W}\right)\leqslant C_\Sigma<\frac{W}{q_W}\log\left(1+\frac{2}{\pi e}\frac{Sq_W\nu_i^2}{\nu^2\sigma_0^2W}\right) \tag{8.40}$$

Figure 8.33 shows curves of the dependence $C_\Sigma = f\frac{P_s}{P_n}$ for various values of n, determined by the upper boundary of formulas (8.40). It follows from the figure that with limited peak power, the throughput capacity of a channel depends on the number of signals multiplexed and decreases with increasing n. /257

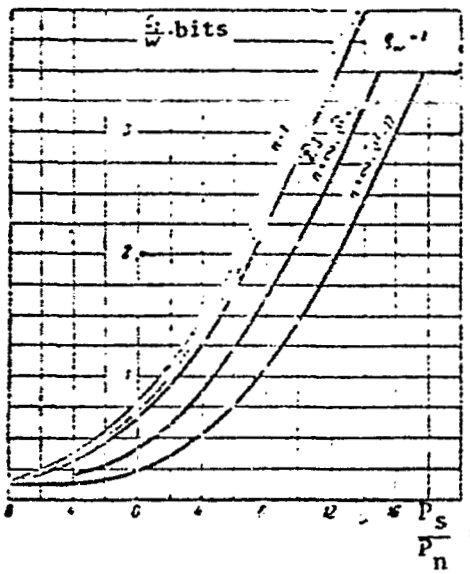


Figure 8.33. Throughput Capacity of Channel with Limited Peak Power with Frequency Division Multiplexing

Let us determine the throughput capacity of a channel with time division multiplexing. With time division multiplexing, the cycling time interval T is divided into n sectors of length T/n. A portion of each sector in time, equal to τ, is used to transmit the signal of one subtrunk. Thus, each of the signals multiplexed has the full power, occupies the entire frequency band of the channel W, but is transmitted only during a portion of the cycle time equal to $T/q_t n$.

As we know, the maximum quantity of information which can be transmitted through a channel is proportional to the product WT. Consequently, the losses in transmission time characteristic for this method of multiplexing can be considered equivalent to losses in frequency band. This analogy will

C-4

be used to determine the throughput capacity of a channel with time division multiplexing.

For a channel with limited mean power, we can assume that each of the partial signals has normal distribution, the transmission interval τ being great enough that the signal can be considered stable. /258

Then, based on (8.32), the throughput capacity of a partial channel

$$C_i = \frac{W}{q_t n} \log\left(1 + \frac{P_s}{P_n}\right),$$

while the throughput capacity of the sum of all channels with identical signals

$$C_\Sigma = \frac{W}{q_t} \log\left(1 + \frac{P_s}{P_n}\right). \qquad (8.41)$$

It follows from (8.41) that the throughput capacity of a channel with limited mean power with time division multiplexing is independent of the number of signals multiplexed and differs from the throughput capacity of the ideal channel only by the factor q_t, considering the presence of the time guard intervals[1].

For a channel with limited peak power, the optimal distribution of signal values is uniform. If in this case each of the signals multiplexed retains its even distribution over time interval τ, based on (8.33) and (8.34) for sufficiently high signal/noise ratios we can write

$$\frac{W}{q_t} \log \frac{2}{\pi e^3} \frac{S}{P_n} < C_\Sigma < \frac{W}{q_t} \log\left(1 + \frac{2}{\pi e} \frac{S}{P_n}\right). \qquad (8.42)$$

It follows from (8.42) that the throughput capacity of a channel with time division multiplexing is independent of the number of signals even with limited peak power. This result is explained by the fact that with time division multiplexing at any moment in time one signal is transmitted and the limitation of peak power does not essentially change the reception conditions, since with sufficiently long time guard intervals, mutual

[1] The general considerations concerning possible dependence of C_Σ on q_w stated in the preceding section are also correct in this case.

interference between signals multiplexed does not occur. The results of calculations using formula (8.41) are shown on Figure 8.34 by the solid lines, while results using formula (8.42) are illustrated by the dotted lines [8.23].

When a channel is multiplexed with respect to waveshape, the signals of all stations are present in the channel simultaneously, each occupying the entire frequency band W. Correspondingly, for each partial signal, interference is represented not only by the fluctuation noise, but also by the signals of the other stations if they are not orthogonal. However, achievement of orthogonality of signals of stations spread over an area requires synchronous operation of the entire network, which is extremely difficult when there are many stations.

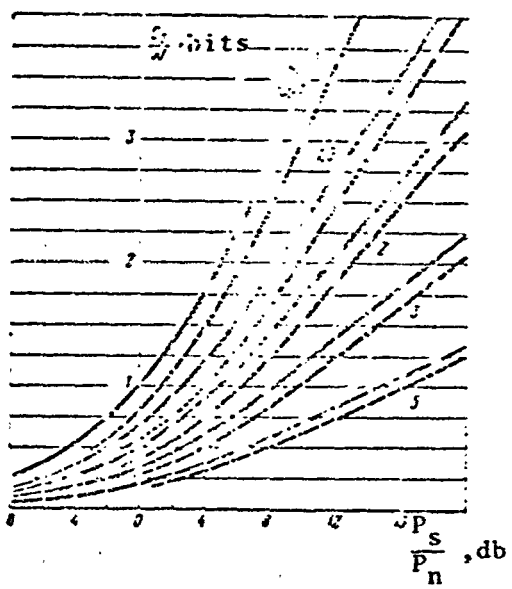


Figure 8.34. Throughput Capacity of a Channel with Time Division Multiplexing

Therefore, we will analyze asynchronous waveshape multiplexing. Since all signals are statistically independent in this case, the mean noise power is equal to the sum of the mean powers of interfering signals and Gaussian noise. The throughput capacity of a partial channel with limited mean power is equal to /259

$$C_1 = W \log \left[1 + \frac{P_1}{(n-1)P_1 + \sigma_0^2 W}\right] \tag{8.43}$$

The throughput capacity of the sum of partial channels will, obviously, be n times greater. Let us express this throughput capacity by means of the parameters of the summary channel:

$$C_\Sigma = nW \log \left[\frac{n\left(1 + \frac{P_S}{P_n}\right)}{(n-1)\frac{P_S}{P_n} + n}\right] \tag{8.44}$$

Figure 8.35 shows a dependence constructed using formulas (8.44). We can see from the figure that the throughput capacity of a channel with waveshape multiplexing depends on the number of stations and decreases with increasing n. /260

For $n \to \infty$, from (8.44), by evaluating an indeterminate form of the type $\infty \cdot 0$, we can easily produce

$$C_\Sigma \to \frac{W}{\ln 2} \frac{P_s/P_n}{1 + (P_s/P_n)},$$

and where $P_s/P_n \gg 1$, the throughput capacity approaches the limit $C_\Sigma \to 1.44\ W$ bits/sec.

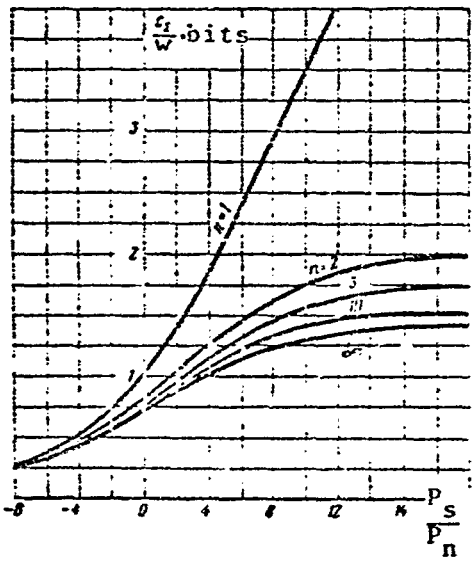


Figure 8.35. Throughput Capacity of a Channel with Limited Mean Power with Asynchronous Form Multiplexing

In a channel with limited peak power, the throughput capacity of the sum of all partial channels expressed through parameters of the summary signal [8.23] is equal to

$$C_\Sigma = nW \log \left[\frac{n\left(1 + \frac{S}{\sqrt{2}P_n}\right)}{n + (n-1)\frac{S}{\sqrt{2}P_n}} \right] \qquad (8.45)$$

The results of calculations using formula (8.45) are presented in Figure 8.36, from which it follows that the limitation in peak power causes no significant decrease in throughput capacity of the channel, which confirms the conclusion drawn earlier that noise-like signals are stable with amplitude limitations. /261

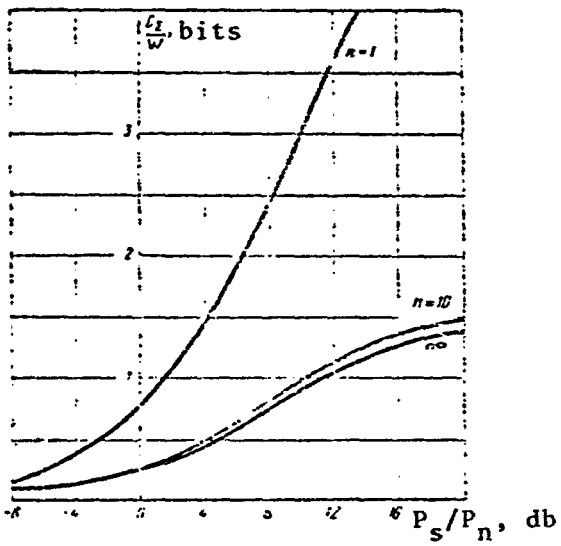


Figure 8.36. Throughput Capacity of a Channel with Limited Peak Power with Asynchronous Waveshape Multiplexing

§ 8.9. Comparison of Multiplexing Methods

Conclusions

The results produced allow us to compare all of the methods of multiplexing studied and select the most effective.

As the effectiveness criterion, we decided earlier to use a certain parameter Z, representing the ratio of the resultant throughput capacity of the communications channel with n signals to the throughput capacity of a channel with 1 signal, i.e.

$$Z = \frac{C_\varepsilon}{C_{n=1}}$$

Generally, this parameter depends not only on the number of stations, but also on the signal/noise ratio in the channel; therefore, various methods of multiplexing should be compared with identical signal/noise ratios. /262

As an example, let us select P_s/P_n = 10 db and, using the formulas presented above, calculate Z.

The results of information estimates based on the formulas of § 8.8 are presented in Figure 8.37, which indicates that the best method is time division multiplexing, the worst method is asynchronous waveshape multiplexing.

Actually, with time division multiplexing the throughput capacity of a channel with any power limitation is independent of the number of signals multiplexed, and when there are no guard bands Z = 1 (top line on Figure 8.37). With frequency division multiplexing, the throughput capacity of a channel with limited mean power is also independent of the number of stations Z = 1. /263 However, with limited peak power, the picture changes sharply and the value of Z decreases with increasing n (dot-dash curves on Figure 8.37).

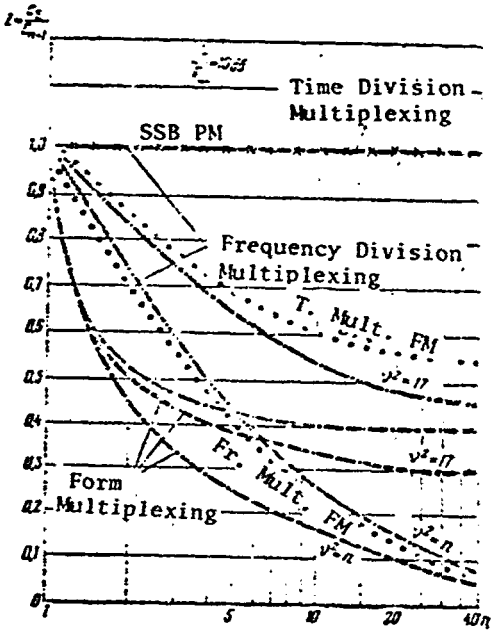


Figure 8.37. Comparison of Multiplexing Methods on the Basis of Effectiveness Criteria Z

With asynchronous waveshape multiplexing (dotted curves), the value of Z decreases with increasing n and reaches its minimum value for a channel with limited peak power.

We should note that with other signal/noise ratios, the values of parameter Z will differ, but the nature of the curves and the result of comparison remain the same.

All of this relates to multiplexing of optimal signals and is correct only with infinitely low probability of error. In order to analyze the situation with actual signals and finite probability of error, let us use the results presented earlier in § 8.3 and § 8.6. On Figure 8.37, the points show a curve constructed on the basis of Figure 8.7 and characterizing the effectiveness of frequency division multiplexing of harmonic signals with frequency modulation with finite probability of error (more precisely, with a fixed signal/noise ratio). Here also, the points show a curve from Figure 8.26 for time division multiplexing of a channel with harmonic signals with frequency modulation with finite probability of error.

Comparison of these curves shows that both in the potential and in the actual respects, the most effective is time division multiplexing. This is followed by frequency division multiplexing and waveshape multiplexing. One exception is the combination of the methods of SSB and PM; it provides the maximum possible effectiveness of multiplexing, but the difficulties in practical realization of this method force us to use simpler systems and, in particular, the system with time division multiplexing.

In conclusion, let us note the possibility of combined multiplexing of signals both as to time and as to frequency. It

is not difficult to see that the signals of ground stations can be made orthogonal by allocating a frequency-time slot to each signal. However, when two ground stations are allocated to the same frequency the problem of synchronization characteristic for systems with time division multiplexing arises. At the same time, the sum of even a small number of ground station signals occupying the same time interval has a high ratio of peak power to mean square power, resulting in ineffective utilization of the relay station power (as is the case with frequency division multiplexing).

Thus, combined multiplexing has the shortcomings of both systems. For these reasons, this method is not used in satellite communications systems.

REFERENCES

8.1. Zyuko, A. G., Pomekhoustoychivost' i effectivnost' sistem svyazi [Interference Stability and Effectiveness of Communications systems], Svyaz'izdat Press, 1963.

8.2. Kalashnikov, N. I., V. L. Bykov, O. S. Krapotin, Radiosvyaz' s pomoshch'yu iskusstvennykh sputnikov Zemli [Radio Communications Using Artificial Earth Satellites], Svyaz' Press, 1964.

8.3. Petrovich, N. T., Ye. F. Kamnev, Voprosy kosmicheskoy radiosvyazi [Problems of Space Radio Communications], Sovetskoye Radio Press, 1965. /264

8.4. Sunde, E. D., "Intermodulation Distortion in Multicarrier FM Systems," IEEE Internat. Conv. Record, 1965, pt. 2, 22-26 March, pp. 130-146.

8.5. Smirnov, V. A., Osnovy radiosvyazi na ul'trakorotkikh volnakh [Principles of Radio Communications at Ultrashort Wavelengths], Svyaz'izdat Press, 1967.

8.6. Bukov, V. L., V. A. Borovkov, S. M. Khomutov, "Energetic Evaluation of Multistation Access with Frequency Compression of FM Signals," Sb. trudov NIIR, 1967, No. 1(46).

8.7. Bukov, V. L., V. A. Borovkov, S. M. Khomutov, "Increasing the Throughput Capacity of Satellite Communications Systems with Multistation Access and Frequency Division Multiplexing," Elektrosvyaz', 1968, No. 1.

8.8. Stewart, I. A., E. A. Huber, "Comparison of Modulation Method for Multiple Access Synchronous Satellite Communications Systems," Electronics Record, 1964, April, pp. 535-546.

8.9. Weaver, C. S., "Comparison of Several Types of Modulation," IRE Transact. Comm. Syst., 1962, Vol. CS-10, No. 1, pp. 96-101.

8.10. Rosen, H. A., "Synchronous Communication Satellites, Wescon 63," Techn. Papers, 1963, Aug., Pt. 6.

8.11. Bukov, V. L., S. M. Khomutov, "The Use of the Combination of SSB and PM in Satellite Communications Systems," Elektrosvyaz', 1967, No. 11.

8.12. Bukov, V. L., V. A. Borovkov, S. M. Khomutov, "Study of a Multistation System with Time Division Multiplexing of FM Signals (Spectral Properties of Signals)," Sb. Trudov NIIR, 1967, No. 4(49).

8.13. Bukov, V. L., V. A. Borovkov, S. M. Khomutov, "Study of Multistation System with Time Division Multiplexing of FM Signals (Problems of Interference Stability)," Sb. Trudov NIIR, 1968, No. 1.

8.14. Tartara, G., "Discontinuous Transmission of Frequency Modulated Waves," Alta Frequenza, 1965, Aug., No. 8, pp. 543-550.

8.15. Kirik, Yu. M., E. Ya. Ryskin, "Interference Stability of FM Signal with Interruptions," Elektrosvyaz', 1968, No. 8.

8.16. Korobkov, L. A., "Device for Multiple Communications," USSR Author's Certificate No. 71601, Byull. izobreteniy, 1949, No. 3.

8.17. Bukov, V. L., B. I. D'yachkov, V. A. Borovkov, S. M. Khomutov, "Power Evaluation of Multistation Access with Time Division Multiplexing of FM Signals," Trudy NIIR, 1968, No. 2.
8.18. Glen, A. B., "Communications System with Code Multiplexing of Channels," Zarubezhnaya radioelektronika, 1965, No. 3.
8.19. Bukhovskiy, M. A., V. S. Rabinovich, "Interference Stability of Multistation Communications System with Complex Signals," Sb. Trudov NIIR, 1967, No. 4(49).
8.20. Bogdanov, L. V., V. L. Bukov, "Influence of Amplitude Limitation on Interference Stability of Communications Systems with Noise Signals," Sb. Trudov NIIR, 1967, No. 2(47).
8.21. Shannon, C., Raboty po teorii informatsii i kibernetike [Works on Information Theory and Cybernetics], Foreign Literature Press, 1963.
8.22. Bukov, V. L., L. V. Bogdanov, "Information Estimate of Methods of Multiplexing of Communications Channels," Sb. Trudov NIIR, 1967, No. 4(49).

CHAPTER 9. PARAMETERS OF THE FM TRANSMISSION SYSTEM OF A COMMUNICATIONS SATELLITE LINK

§ 9.1. Selection of Frequency Deviation /265

The first task in determining the parameters of the high frequency section of a communications link using frequency modulation is selection of the frequency deviation. The less the deviation in frequency, the simpler it is to develop apparatus, the less the nonlinear distortions of the messages transmitted. However, the greater the frequency deviation, the easier it is to achieve the signal to noise ratio required at the output of the line due to the increasing FM gain (see Chapter 7). Due to the power limitations of satellite communications lines, the second requirement usually predominates -- maximum increases in frequency deviation. This tendency is limited by arrival at the so-called threshold of FM interference stability. This threshold is manifested in the fact that as the signal to noise ratio decreases to a certain value, the FM gain drops sharply (see Figure 9.1). Furthermore, when the threshold is reached, the nature of the noise at the output of the receiver changes: in addition to the "smooth" noise with normal distribution and parabolic power spectrum, short, powerful pulses with almost even spectrum appear. The effect of the threshold and methods of combatting it will be described more completely in the second volume of this monograph (see also 9.1). At present, we are interested only in the fact that with increasing modulation index

$$m = \Delta f_d / F_B,$$

i.e., with increasing frequency deviation Δf_d, the threshold begins to occur at higher signal powers. Below the threshold, noise increases very rapidly. Therefore, a given ratio of signal to noise, fixed in planning a radio link (for example $H_{md1} = (P_s/F_n)_{out1}$ corresponds to an optimal modulation index (for example, m_2), for which the necessary value of H_{md1} is achieved /266 in the area of the threshold. With a lower modulation index (for example, m_1), the FM gain will be less and, with a given noise factor of the receiver device, a higher signal to noise ratio is required (H_{d1} on Figure 9.1). With a higher modulation index (m_3), reception will occur below the threshold, also requiring

greater signal power in order to bring the signal-to-noise ratio at the input of the receiver to H_{d3}. For certain types of messages (for example, telegraph messages), pulse noises are particularly dangerous; therefore, operation below the threshold is not permissible, and with index n_3, the signal-to-noise ratio must be increased to H_{d4}. The threshold curves presented in Figure 9.1 were constructed by calculation for a standard FM signal demodulator, consisting of an ideal limiter and an ideal frequency detector (differentiator) according to data from [9.2]. In the portion of these curves above the threshold, the potential interference stability of the method of frequency modulation is fully realized [9.3]. The area of the threshold is calculated by analysis of the derivative of the phase of the sum of signal and noise, i.e., is correct for the standard system of frequency demodulation just mentioned.

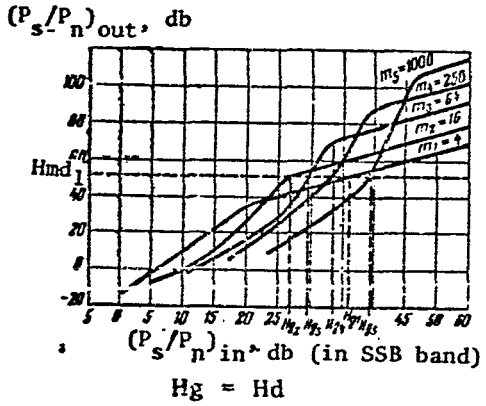


Figure 9.1. Threshold Characteristics of FM Signal Receiver

In satellite communications systems, various methods are usually used to increase the interference stability of FM reception, shifting the threshold into the area of lower values of signal-to-noise ratio [9.1]. These methods are based on the fact that the bands of frequencies occupied by the FM signal need not necessarily be fully transmitted by the receiver at each moment in time. D. V. Agenev [9.4] has introduced the concept of the active band of the spectrum, or more simply the active spectrum. The active band of function f(t) at moment t - τ refers to the minimum possible band width from $\omega_1(t)$ to $\omega_2(t)$ for which the function

$$f_\Delta(t-\tau)=\frac{1}{\pi}\int_{\omega_1(t)}^{\omega_2(t)} S_t(\omega)\,e^{i\omega(t-\tau)}\,d\omega$$

is similar to function f(t - τ) with fixed accuracy and according to a criterion determined in advance (for example, nonlinear distortion). Here τ is the time of passage through a filter /267

with band width $\omega_1(t)-\omega_2(t)$ at moment t,

$$S_t(\omega) = \int_{t_0}^{t} f(t) e^{-i\omega t} dt$$

is the complex instantaneous spectrum of function f(t).

For an FM signal, the active band is placed such that the instantaneous frequency of the signal $\omega(t)$ is at the center of the active band, while the width of the active band is determined by the rate of change of frequency, i.e., the frequency of the modulating process, and is independent of the frequency deviation. Thus, reception of an FM signal can be conducted using a filtering system with a pass band, the central frequency of which tracks the instantaneous frequency of the FM signal, while the width (band) at each moment is significantly narrower than the total spectrum of the FM signal, approaching the band width of an AM receiver (i.e., $2F_B$), more precisely, the more precisely the tracking process is performed. When a signal and fluctuation noise are received, the tracking receiver decreases the noise power entering the frequency detector, since the fluctuation noise is evenly distributed over the entire band, and the process of tracking the filter does not annul the gain achieved by the narrower band width of the filter. At the same time, the power of the FM signal at the input of the frequency detector itself remains unchanged. Therefore, the threshold effect associated with the bursts of noise reaching the level of the incoming signal, occurs at a later time when the signal level is lower.

In the area above the threshold, naturally, with tracking reception the gain due to the narrow filter at the input of the frequency detector is 0, since (and if) a modulating frequency filter with limiting frequency F_B is installed at the output of the receiver; this corresponds to the fact of realization of the potential gain of FM by a standard receiver mentioned earlier.

Without entering into a number of fine points of the operation of various circuits shifting the threshold at present, let us note only that analysis of threshold conditions for these systems is quite complex and has been performed only very approximately as yet. The capabilities of known demodulator circuits differ: for various modulation indices, the best results are achieved using various types of circuits. Finally, we should note that the circuits of interference-stable demodulators are in a stage of rapid development.

Planning prospective communications system links, particularly when decisions concerning the distribution of the spectrum

of radio frequencies, difficult to adjust later, are to be made, it is sensible in many cases to base discussions not on the capabilities of actually available circuits, but rather on the theoretical limit, which cannot be exceeded by actual demodulators. This limit is defined in [9.5]. The definition is based on the idea, suggested by A. G. Zyuko [9.6], that the limiting position of the threshold can be determined from the condition of equality of throughput capacities of the HF line (C_1) and of /268 the same line, based on modulating spectrum (based on message channel) for the area above the threshold (C_2). In the area above the threshold, $C_1 > C_2$, i.e., with frequency modulation less than the entire throughput capacity of the channel as used to transmit the message. As the threshold is approached, the throughput capacities C_1 and C_2 drop, C_1 decreasing more rapidly. Since it is obvious that always $C_1 \geq C_2$, the condition $C_1 = C_2$ is the limit, below which the formula for interference stability of FM for the area above the threshold cannot be correct, i.e., the limit for left shift of the threshold point.

To calculate the value of C_1, it is simplest to take as a basis the most general relationship for throughput capacity

$$C_1 = 2W[H(y) - H_x(y)],$$

where H(y) is the entropy of the distribution of the signal receiver (per character), $H_x(y)$ is the conditional entropy (per character) of the signal received (assuming that signal x was transmitted), W is the frequency band of the high frequency line. According to the usual conditions, we will consider the noise additive, i.e., that it is simply added to the signal with no mutual influence between signal and noise. Let us also consider the noise to have normal distribution and zero mean value.

Then [9.7]

$$C_1 = 2W[H(y) - H(n)]. \qquad (9.1)$$

where H(n) is the noise entropy, where

$$H(n) = \ln \sqrt{2\pi e}\,\sigma,$$

σ^2 is the dispersion of the noise.

In order to determine H(y), we must find the distribution function of the sum of the sinusoidal signal and normal noise. The distribution of the phase in this case is assumed even (i.e., the modulating function is, for example, triangular). The distribution function $p_1(y)$ can be obtained by inverse Fourier transformation from the product of the characteristic functions of the components [9.5], [9.8].

The definition of entropy

$$H(y) = -\int_{-\infty}^{\infty} p_1(y) \ln p_1(y)\, dy$$

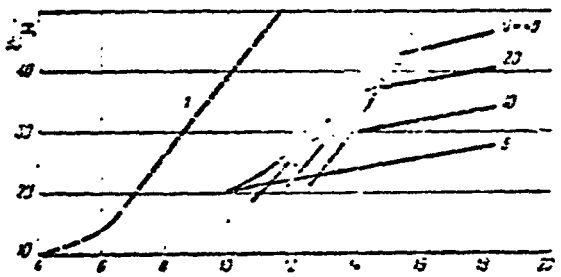
can be used directly for numerical calculation of the throughput capacity of the HF channel C_1/W [9.5]. Approximately the same results are produced for frequency modulation of the signal by the normal process.

In order to calculate the information transmission rate through the LF line C_2, we can use the same initial relationship (9.1), if we consider that right up to the moment of onset of the threshold, the same statistical properties of signal and noise are retained as in the area above the threshold with a standard demodulator. For this case, the distribution of noise is near normal [9.8]. The distribution of signal at the output of the receiver with low distortions in the line corresponds to the distribution of the modulating function. Therefore, in accordance with the distribution of phase used in calculation of C_1, we will consider that the information (LF) signal has even distribution, which yields the maximum rate of information transmission C_2 with a limitation on peak power (corresponding to the limiting value of frequency deviation). /269

We find under these conditions that

$$C_2 = F \ln 0{,}7 \quad (P_s/P_n)_{out}$$

Very similar results were produced using a model of a message in the form of a sine wave with random phase. This model is convenient for studying the conditions of transmission of television signals.



$(P_s/P_n)_{in}$, db (in SSB band)

Figure 9.2. Determination of Limiting Position of Threshold in Communications Lines with FM (with Sinusoidal Modulating Signal)

The conditions of transmission of multichannel telephone signals with frequency division multiplexing can be more easily understood if we consider the message a normal process.

In this case[1]

$$C_{2n} = F\ln(P_s/P_n)_{out}.$$

Finally, equating C_1 and C_2, we produce curves 2, 3 and 4 (Figure 9.2) of the limiting interference stability of FM reception.

These curves depend on the ratio ν of the band widths of the HF and LF lines and, consequently, on modulation index m. The intersection of the curves with the lines for operation beyond the threshold indicates the limiting position of threshold points. Curve 1 illustrates the limit reached by an ideal receiver (according to Shannon) where $\nu \to \infty$, which allows us to estimate the losses in throughput capacity, unavoidable in principle, resulting from the use of FM. /270

Figure 9.3 shows the position of the envelopes of the threshold points, both limiting[2] and those constructed by averaging of results experimentally produced by various authors for demodulators with frequency feedback (FFB). The results achieved using other tracking reception systems are similar to those illustrated.

Let us use an example to show how Figure 9.3 is used to select the frequency deviation for a satellite communications link. Suppose we must transmit a television signal according to the ICCR recommendations. In this case, as was shown in Chapter 6, the ratio of the power of the sinusoidal signal (with amplitude equal to the amplitude of the video signal plus synchronization pulses)

[1]It should not be thought that the absence of the factor 0.7 makes $C_{2n} > C_2$; due to the greater peak factor of the normal process, the mean square value of the signal P_s with limited peak power is considerably less than with even distribution.

[2]The model of the message used is a sine wave with random phase.

to the power of noise at the output of the frequency detector should be 31.2 db (shown by dotted line on Figure 9.3). If the receiver device uses a standard demodulator (curve 3), the best modulation index will be about 3.3, the frequency deviation will be 3.3·6 = 19.8 MHz. The necessary signal-to-noise ratio at the input of the receiver system is about 10 db with a band width of 40 MHz.

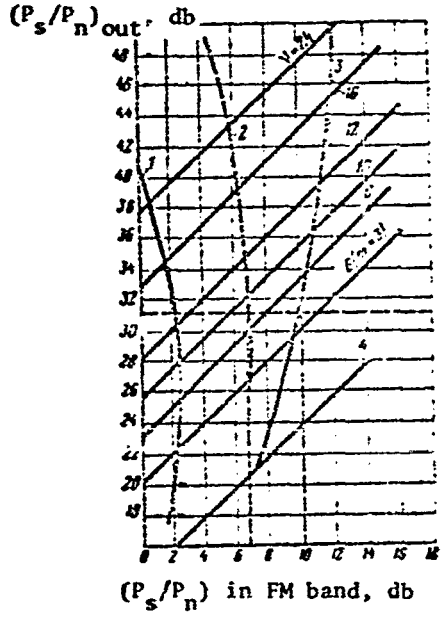


Figure 9.3. Envelopes of Threshold Points (Sinusoidal Modulating Signal): 1, Limiting Interference Stability of Demodulator (Theoretical Data); 2, Demodulator with FFB (Experimental Data); 3, Standard Demodulator

The use of the limiting interference stable demodulator results in the prospective value of m = 6.5, for a frequency deviation of 6.5 x x 6 = 40 MHz. This requires a ratio of only 1.5 db at the input with a band width of about 80 MHz, i.e., with the same noise temperature of the receiver (8.5 + 10 log 80/40) = /271 = 11.5 db less than when a standard demodulator is used. The band width of 80 MHz is the extreme value which could ever be required using FM for transmission of television. Finally, for reception using a demodulator with FFB, $m_{opt} = 4.5$, or, since a certain threshold reserve is desirable, $m_{opt} = 4$.

As we see, the gain due to the use of interference-stable demodulators is significant. However, this gain is achieved only with proper selection of the parameters of the FM signal. If the mode of the line is selected assuming a standard demodulator, the use of interference stable demodulators does not improve the quality of reception; it only creates a threshold reserve, useful at times when the power characteristics of the line deteriorate in comparison with the design characteristics.

Up to now, the selection of deviation has been based on a single criterion -- achievement of the minimum necessary signal level at the input of the receiver, i.e., the maximum savings of transmitter power for the satellite relay station. In actual

planning, it may be necessary to decrease the frequency deviation to decrease distortions of the FM signal in electronic elements of the line (see §§ 9.2, 9.3), to increase selectivity, i.e., to decrease interference between radio signals with adjacent frequencies or, finally, and this will be very important in the future, to save room in the radio spectrum, i.e., to increase the number of signals which can be placed in a given frequency band.

§ 9.2. Selection of Parameters of HF Line from the Point of View of Nonlinear Distortions

Distribution of nonlinear distortions between line elements. The necessity of fulfilling the recommendations for nonlinear distortions of messages transmitted is a determining condition for the selection of the most important characteristics of a line in the case of frequency modulation. In the case of transmission of television, these distortions are normalized directly, while in the case of multichannel telephony they lead to crosstalk noise.

The overall recommendations for any type of distortion, fixed for the entire line, should be distributed among line elements, i.e., primarily among the two ground stations and the satellite relay station, then among the component parts of the equipment within the limits of each of these elements. The procedure for distribution of distortions among line elements should, of course, consider the capabilities of each element, resulting from the current state of development of equipment, as well as the more basic aspect of the problem -- the functions of the elements.

Thus, for example, the distribution of the tolerance for nonlinear distortions among elements of a ground station receiver should place the strictest possible requirements on the frequency detector, since linearization of the FD simultaneously decreases the influence of interference from the neighboring channel on the receiver; it should be kept in mind that an increase in linearity usually is accompanied by a decrease in steepness of the FD, increasing the requirements to suppression of amplitude modulation in the limiter. Conversely, the IF amplifier of the receiver should be allowed the maximum possible distortion norm, since this allows the IF amp to be made with the narrowest possible band width and the steepest slope, increasing the interference stability of reception. Thus, the operation of allocation of distortions among elements requires that the planner of a radio line understand the specific properties of the elements well. /272

In distributing the distortion norm among elements of the line, the problem arises of the addition of distortions arising in various elements. The most unfavorable is the case of linear (arithmetic) addition of distortions. Then each link of the line,

consisting of m identical elements, should have distortions not over

$$n_{\tau} = \frac{n_{\tau\Sigma}}{m}.$$

where $n_{T\Sigma}$ is the total distortion in a line of m sections. With nonidentical distortions in the elements

$$n_{\tau\Sigma} = \sum_{m=1}^{m} n_{\tau m}.$$

In practice, the most favorable possibility must be considered as addition of distortions with respect to power; addition of distortions arising in different sectors of a multielement line independently, for various reasons and in various phases, can be reduced to this case. The permissible distortions in one element in this case

$$n_{\tau m} = \frac{n_{\tau\Sigma}}{\sqrt{m}}$$

or with various distortions

$$n_{\tau\Sigma} = \sqrt{\sum_{m=1}^{m} n_{\tau m}^2}.$$

i.e., quadratic addition of distortions occurs. The quadratic rule is expedient for use upon addition of transient noises arising in various sections of the line upon transmission of multichannel telephony with frequency division multiplexing.

Finally, in certain cases the best results are produced by an intermediate rule of addition, the "3/2 power" rule, i.e., /273

$$n_{\tau} = \frac{n_{\tau\Sigma}}{m^{\frac{2}{3}}}.$$

ICCR Recommendation 421 considers this method of addition most favorable for calculation of nonlinear distortions in a long black and white television transmission line containing over three sections, forming a standard circuit. It is considered suitable to use this rule for addition of nonlinear distortions in satellite communications line elements. An exception should be made in the case of addition of frequency-dependent nonlinear distortions, arising in video amplifiers, frequency detectors, frequency modulators with frequency-dependent message distortions arising in the radio line due to the non-ideal nature of its linear characteristics. Since the products of these two types of distortion usually differ in phase by 90°, in calculating the resultant harmonic factor, one should use the quadratic rule. In calculating the value of $n_{T\Sigma}$, distortions in the FM link can be ignored, since only crosstalk phase modulation of the two-component selective signal arises in this link, caused by the 90° rotation of the products of distortions.

In planning satellite communications lines, we must be able to relate the various nonlinear distortion factors used in normalization and calculation of systems. Thus, in specifications for nonlinear distortions in television systems, the value n_T is used, the sense of which is explained in Chapter 6.

This index gives the direct value of non-uniformity of the differential characteristic, usually used to estimate distortions in frequency detectors, modulators and video amplifiers. The factor n_T is related to the best known indicator of distortion -- the harmonic factor (in the case of frequency-independent distortion) as follows:

-- with quadratic distortions

$$n_{T2} = 8k_{h2}$$

-- with cubic distortions

$$n_{T3} = 12k_{h3}$$

Thus, the index of distortions used in television has much more significance for this purpose than the harmonic factor; the popular opinion that great distortions are permissible in television (n_T = 20%!) is not very well founded.

As we have already pointed out, as an FM signal passes through linear circuits with nonlinear phase-frequency and

non-uniform amplitude-frequency responses, nonlinear distortions of the message arise, the products of distortions being proportional to the modulation frequency [9.1]. In this case, we can compare the value of n_T with the value of the harmonic factor, measured at frequency F_B of the high frequency filler signal (about 1.2 MHz) with the amplitude of the test signal at this frequency equal to the nominal level of the video signal. Since the relationship of the frequencies of the two components of the test signal used for measurement of n_T is quite high, the entire spectrum of the products of distortions, located in the region of the HF component and creating its cross modulation, is narrow. Therefore, the frequency of any of these products may be assumed equal to frequency F_B. As we have already noted, due to the rotation of the phase of these products of distortions, the value of n_T in FM links can be considered equal to 0. /274

Influence of linear predistortions. Let us study the influence of predistortions of the frequency characteristic ordinarily used on radio relay and satellite lines with FM on the amplitude of the video signal, noise and nonlinear distortions.

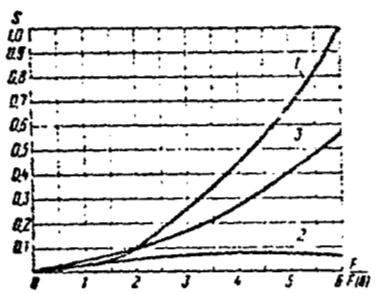


Figure 9.4. Nature of Noise Power Spectrum at Output of FM Receiver: 1, Without Predistortions; 2, With Predistortions Used in Radio Broadcasting; 3, With Predistortions for Transmission of Television over Radio Relay Lines

We know that the fluctuating noise in a high frequency radio relay line creates noise at the output of an FM receiver, the spectral power density of which increases at the high frequencies of the band. Therefore, the introduction of predistortion usually produces a gain in interference stability of the line, which can be estimated from the integral relationship of signal and noise powers at the output of the FM receiver. Thus, in the case of radio broadcases with FM, the gain reaches 10 db [9.10], while the change in noise power spectrum is illustrated by curve 2 on Figure 9.4.

When black and white television is transmitted over radio relay or satellite communiations links with FM, the predistortion curve is standardized, involving some increase in the high frequencies plus significant attenuation of the low frequencies of the modulating function spectrum [9.11]. Therefore, the spectral density of noise power at the output of the receiver after the restoration circuit not only decreases, but also

increases in the low frequency section. The integrated noise power is still somewhat decreased -- by 2 db and 2.2 db for measurements in the 5 MHz and 6 MHz bands respectively. However, noises with /275
television transmission are always measured with a weighting filter, considering the rapid drop in visual noticeability of noises with frequency. Then the noise spectrum at the output of the FM receiver follows curve 1 on Figure 9.5, or with predistortions -- curve 2. The drop in the higher frequencies in the weighting filter, naturally, decreases the effect of the noise gain in the upper section of the spectrum due to predistortions, while the loss at the lower frequencies is almost fully retained; therefore, the integral noise in this case is decreased by predistortions very little (0.2-0.4 db), i.e., remains practically unchanged.

This conclusion was found by numerical integration of the spectral noise density curve at the output of the FM receiver in the area over the threshold, multiplied by the transmission factor of the restoring and weighting filters. Comparison is performed with the case when there is no restoring filter, i.e., predistortions are not introduced, it being assumed that after connection into the predistorting and restoring filter line, the level of the modulating signal at the input of the line does not change, so that the signal at the output of the line remains unchanged. This condition has not been fully founded.

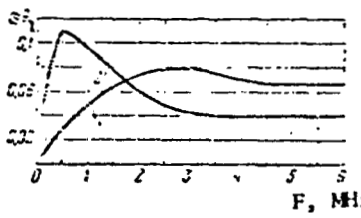


Figure 9.5. Noise Spectrum at Output of FM Receiver with Weighting Filter: 1, Without Predistortions; 2, With Predistortions

Actually, we know that the main portion of the power of a television signal is concentrated in the low and middle frequencies, where the predistorting circuit produces significant attenuation. Therefore, it can be assumed that the amplitude of the television signal will be decreased by the predistorting circuit. On the other hand, the synchronizing and damping pulses, like the pulses of the video signal itself, are differentiated in the predistorting circuit. If the leading edges of the pulses are sufficiently steep, the amplitude of excursions formed after passage through the predistorting circuit may be even greater than the amplitude of the pulses at the input of the circuit. However, if the steepness of the leading edges is limited to those values standardized for television signal synchronization mixtures, the amplitude of excursions is not over the amplitude of the input pulses. With preliminary (i.e., before predistortion) transmission of pulses with arbitrarily steep leading edges through multiple-element RC low frequency filters with limiting frequency 5-6 MHz, the excursions resulting from the

predistorting circuit also will not exceed the amplitude of pulses at the input of the circuit. Furthermore, a number of observations have been made of actual television signals passing through predistorting circuits after passing through a line with limited band width. The overall conclusion of all observations is as follows: the amplitude of a video signal which has passed through a predistorting circuit depends strongly on the width and shape of the frequency characteristic of the line through which it has passed. There is reason to believe that under actual operating conditions of radio relay or satellite communications lines with FM, the input of these lines will receive a television signal containing no components with frequencies higher than 6 MHz. Under these conditions, when radio relay and satellite communications lines with FM are planned, one should assume that the amplitude of the television signal will not change due to predistortion. /276

Let us now study the influence of linear predistortion on nonlinear distortions of messages.

When predistortions are introduced, the nonlinear distortions, estimated on the basis of the harmonic factor, are decreased in accordance with the frequency characteristic of the restoration circuit. When cross modulation methods are used in the measurements, the entire spectrum of the modulated video frequency (1.2 MHz) oscillation, as was noted above, is narrow, and the relationship between the carrier and the side bands is not disrupted by the restoration circuit. Therefore, the influence of the predistorting circuit results entirely from the attenuation of the low frequency test signal which it produces. This attenuation is approximately 8 db (2.5 times with respect to voltage), while the amplitude of the two-component television test signal, after passing through the predistorting circuit, is approximately half the nominal amplitude of the video signal. Consequently, quadratic nonlinear distortions are decreased by approximately 2.5 times, cubic distortions -- by six times. This allows us to reduce the requirements for linearity of the television transmission line; on the other hand, with unchanged linearity, predistortions make it possible to increase the amplitude of the video signal by approximately two times and correspondingly increase the frequency deviation if the applicable norms for nonlinear distortions are observed.

Of course, in practice the increase in frequency deviation may be limited by other factors. In this connection, we must mention the earlier onset of the FM gain threshold when the instantaneous frequency moves outside the transmission band of the high frequency line [9.12]. This criterion should be checked for the full amplitude of the television signal.

Finally, decreasing the level of the constant component of the video signal almost to zero, predistortions decrease the

degree of utilization of the modulation characteristics and therefore decrease distortions.

Nonlinear distortions of color television signals and multichannel telephone signals. We must establish the relationship between characteristics for color television and the rigidly normalized (particularly in the NTSC system) "differential phase" and "differential gain" distortions and the parameters of the high frequency line. /277

In agreement with [9.9], let us define the "differential phase" as the change in phase of the color subcarrier signal with a change in luminance signal from the level of black to the level of white. With this definition, the dependence of "differential phase" on brightness signal level can be sought as an additional phase angle, produced by the component with the frequency of the color subcarrier.

The sum of the sinusoidal luminance signal and the unmodulated color subcarrier can be written as

$$S(t)=U_0(\sin\Omega t+k\sin\omega_c t), \tag{9.2}$$

where ω_c is the frequency of the color subcarrier

$$\omega_c\gg\Omega \text{ and } k\ll 1.$$

As the high frequency FM signal is transmitted, the change in frequency of which is determined by this expression, through a radio line with nonlinear phase characteristic, the modulating signal experiences nonlinear distortions, resulting in the appearance of an additional phase shift of the subcarrier, in contrast to the nonlinearity of the video line, causing only a change in the amplitude of the main components of the modulating signal and appearance of new harmonics. Actually, the expression determining the modulating signal at the output of the frequency detector, according to [9.3], is

$$S_0=S+\psi_2\Delta\omega_d(S^2)'+\psi_3\Delta\omega_d^2(S^3)', \tag{9.3}$$

where $\Delta\omega_d$ is the frequency deviation of the high frequency signal;

ψ_2 and ψ_3 are the coefficients of the exponential polynomial approximating the phase characteristic;

S(t) is the message transmitted.

Substituting (9.2) into (9.3) and anlyzing the expression produced, we see that the value of the "differential phase" is determined by the quadratic component with frequency ω_c and is (where k << 1)

$$\theta_3 = \arctan \frac{3}{2}\psi_3 \Delta\omega_d^2 \omega_c = \arctan 2k_{h3}, \qquad (9.4)$$

$$\theta_2 = \arctan 2k_{h2}, \qquad (9.5)$$

where k_{h2} and k_{h3} are the coefficients of nonlinear distortions in the second and third harmonics respectively, measured at the frequency of the color subcarrier under the condition that the amplitude of the test signal is equal to the amplitude of the video signal. It follows from (9.4) that the value of "differential phase" is independent of frequency of the low frequency oscillation; consequently, we can replace the low frequency sinusoidal signal with a low frequency signal of more complex form, in particular a sawtooth signal. /278

The term "differential gain" refers to the dependence of gain of the line at the frequency of the color subcarrier on the level of the luminance signal, i.e., the "differential gain" characterizes the nonlinear distortions of the line at the frequency of the color subcarrier. The differential phase and differential gain are estimated using a test signal similar to the signal received for estimation of linear distortions of the black and white image, with the single difference that the frequency of the sinusoidal filler is selected equal to the frequency of the color subcarrier. The value of differential gain in this case is determined from the change in amplitude of the sinusoidal filler, i.e., is determined like the nonlinear distortion factor n_T. The relationship between factor n_T and nonlinear distortion factor for the second and third harmonics was illustrated above. The use of predistortions decreases the amplitude of the low frequency saw tooth signal, causing a reduction in differential phase and differential gain.

Let us now go over to the case when an FM signal is used to transmit multichannel telephony with frequency division multiplexing. Nonlinear distortions in this case are dangerous, not primarily as distortions of each telephone conversation, but as a factor causing the formation of unintelligible crosstalk interference between channels. This interference is also present during pauses in transmission over the channel in question, and in fact is particularly noticeable in this case. This type of interference is generally calculated and measured under conditions when all channels are loaded except the tested channel in which there is no load.

Calculation of crosstalk noise during transmission of multichannel telephony with frequency division multiplexing cannot be performed satisfactorily if the modulating function is represented in deterministic form. The most suitable message model in this case is a stable random process with normal distribution, with a spectrum distributed evenly over the band of frequencies from F_1 to F_2 -- the band of frequencies of the multichannel message. Let us present a relationship which is convenient for calculations [9.3]:

$$P_{1k} = e^{2b_k} \frac{\Delta F_k k_{n1}^2}{F_2 - F_1} \left[4e^{4b_m} y_2(\sigma_k) k_{g2}^2 + 24e^{6b_m} y_3(\sigma_k) k_{n3}^3\right] \cdot 10^9 \text{ pw}, \qquad (9.6)$$

where b_k is the measured level of one channel, nep;

b_m is the difference between the mean power of the multichannel message and the power of the measurement signal in one channel [9.13];

$\sigma_k = \frac{F_\nu - F_1}{F_2 - F_1}$ is the relative frequency of the channel;

$y_2(\sigma_k)$, $y_3(\sigma_k)$ are functions characterizing the spectral density of the products of distortions (Figure 9.6); /279

$k_p = 0.75$ is the psophometric factor;

ΔF_k is the band of a channel;

k_{n2} and k_{n3} are the harmonic factors in the second and third harmonics, measured at the measurement level of one telephone channel.

Formula (9.6) can be used to calculate noise both in those elements of the link where the noises are frequency dependent (modulator, demodulator, video amplifiers) and in those elements where frequency-dependent distortions of the FM signal arise (RF and IF amplifiers). In the latter case, the values of k_{h2} and k_{h3} should be measured at frequencies $F_k/2$ and $F_k/3$ respectively, where F_k is the middle frequency of the channel for which noise power is being determined. In planning communications systems, formula (9.6) must be used in reverse order -- determining the permissible values of harmonic factors on the basis of required noise per channel P_{nk}. This requires that the relationship between k_{h2} and k_{h3} be assigned in advance on the basis of planning experience.

When formula (9.6) is used, extremely low permissible values of k_{h2} and k_{h3} are produced. Actually, satisfactory multichannel

telephony transmission does require extremely low distortions. However, one should not think that these requirements differ radically from the conditions of transmission of television, particularly color television. The fact is that these low values relate to the level of one channel, whereas the peak level of the multichannel message is many times greater -- the number of times being

$$K_p = \frac{1}{\sqrt{2}} \cdot 10^{1/20(b_m(db)+\nu_N)},$$

where ν_N is the peak factor of the multichannel (N-channel) message (generally 9-13 db, [9.15]). Consequently, the distortions permissible with the peak deviation of a multichannel message are K_p (for k_{h2}) and K_p^2 (for k_{h3}) times greater. If we also consider that the frequency at which k_{h2} and k_{h3} are estimated is generally lower than the frequency of the color subcarrier, as well as the fact that very rigid limitations are placed on differential phase, usually the conditions for transmission of color television programs are even more rigid. /280

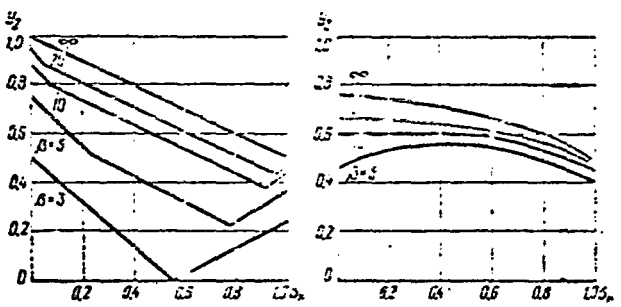


Figure 9.6 Spectral Density of Products of Square and Cubic Distortions

Influence of phase-frequency and amplitude-frequency responses. In satellite communications lines, in contrast to radio relay lines, due to the limited capabilities for spatial selection and due to the power limitations of the radio lines used, there is a general tendency toward most complete possible utilization of the spectrum, i.e., first of all the frequency deviation is made as close as possible to half the transmission band width and, secondly, the protective frequency interval between neighboring trunks is made as small as possible. The first fact requires the use of effective multiple-element

correctors of the phase-frequency and, possibly, the amplitude-frequency responses of the line. Reduction of the guard bands makes it necessary to use multiple-element filters with steep slopes. All of this taken together gives the phase-frequency responses a fairly flat shape with many, slight fluctuations within the limits of the transmission band and a sharp bend at the edges of the band.

These responses are usually unsatisfactorily approximated by cubic polynomials, as was attempted in the classical works on radio relay lines [9.3], [9.13]. Therefore, a tendency is observed to calculate noise considering the phase characteristic components of fifth and even seventh order. The calculation formulas are not presented here, due to space limitations (for example, see [9.14]).

In conclusion, we must establish the relationship between distortion factors and parameters of the phase-frequency and amplitude-frequency responses of the HF line causing these distortions. Let us indicate this relationship only in the simplest case, when analysis is conducted with the so-called quasi-stable approximation, i.e., when all transient processes are ignored; the voltage at the output of the circuit is calculated as if in the stable mode, is considered equal to the product of the FM emf times the static value of the transmission factor of the circuit at a frequency equal to the instantaneous frequency of the emf. Then, the distortions of the message are determined only by the nonlinearity of the phase-frequency response. If the phase-frequency response is fixed in the form of a cubic polynomial

$$\varphi(\omega_\Delta) = \varphi_1\omega_\Delta + \varphi_2\omega_\Delta^2 - \varphi_3\omega_\Delta^2,$$

where ω_Δ is the detuning relative to the center of the transmission band of the amplifier, then /281

$$k_{g2} = \phi_2 \Delta\omega_d \Omega$$

$$k_{3g} = \frac{3}{4}\ \phi_3 \Delta\omega_d^2 \Omega,$$

where $\Delta\omega_d$ is the amplitude of the sinusoidal frequency deviation,

Ω is the modulating frequency.

It is frequently most convenient to use values of delay time asymmetry

$$\delta\tau_m = 2\psi_2\Delta\omega_d$$

and mean unevenness of delay time

$$\Delta\tau_m = 3\psi_2\Delta\omega_d^2,$$

i.e.,

$$k_{g2} = \frac{1}{2}\delta\tau_m\Omega; \qquad k_{g3} = \frac{1}{4}\Delta\tau_m\Omega.$$

The quasi-stable approximation always yields an elevated value of noise [9.1]; therefore, it is usable for initial planning of a line. It should be considered that nonlinear noises arising in real lines are dependent not only on the phase characteristic, but also on the amplitude-frequency response of the HF system. This is particularly significant in the case of systems with large numbers of channels with relatively slight deviation per channel, when effective deviation of frequency for the multichannel message is less than the upper limit of the frequency spectrum of the message [9.15].

In more complex cases, distortions can be calculated by a spectral method; the accuracy of calculations in this case can be arbitrarily high with corresponding increases in expenditure of time [9.1]. Calculations are facilitated by the use of computers.

In determining the recommendations for linearity of phase characteristics of the line, one should consider the possibility of improving these characteristics by means of phase correctors. For example, modern concentrated-selection filters, installed in the intermediate frequency system, with a selectivity as indicated at the end of the next section and a transmission band of 30-40 MHz, have group delay time non-uniformity of 50-100 msec. This quantity can be decreased using an ordinary phase corrector by 2 to 3 times, by 7 to 10 times by using a large number of carefully selected links with nonidentical parameters. The degree of correction is generally limited by the instability of the characteristics of the line corrected. Incidentally, this difficulty is also avoided by rapidly adjustable or even automatically adjusted correctors.

§ 9.3. Factors Involved in the Selection of HF Link Parameters for Transmission of Telephony with Multistation Access /282

First of all, let us indicate some of the factors associated with the selection of the optimal frequency deviation for transmission of

one FM signal modulated by multichannel telephony with frequency division multiplexing. For this purpose, we can use the curves of Figure 9.3, but with some significant adjustments. First of all, it should be kept in mind that in the different channels of the multichannel telephony message, the threshold arrives at different times; it first arrives in the lowest frequency channel, since the noise above threshold is least in this channel and the addition of the threshold noise, which has an approximately uniform spectrum, is felt earlier than in the highest channel. To a certain extent, this phenomenon can be eliminated by linear predistortion of the message at the input of the modulator, decreasing the level of modulation in the lower channel and increasing the level in the upper channel. We should also consider that the onset of the threshold, so dangerous for transmission of pulse signals, is not as dangerous for the case of multichannel telephony, since the threshold noise, which is pulse noise over the entire band of the multichannel message (group band) is heard as a simple addition to the fluctuating noise in each individual channel. Nevertheless, a sharp increase in noise below the threshold makes it optimal as before to operate above the threshold.

Considering all of these considerations, let us limit ourselves to selection of the mode with respect to the onset of the threshold in the group line of a multichannel message. This message can be well imitated by a normal random process. The envelopes of the threshold values, constructed according to the principles outlined in § 9.1 for a normal modulating process, differ little from the envelopes shown in Figure 9.3. They are merely shifted slightly to the left, and in practice this difference can be ignored. Since we have agreed to determine the mode for the group line, while the recommendations for the signal/noise ratio is fixed for an individual telephone channel, we must go over to the recommendations for signal/noise ratio for the group. The line operates near the threshold, but the relationship for spectral noise density in the area over the threshold remains correct:

$$p_n = kF^2.$$

The power of noise in the upper channel of the multichannel system, to which it is convenient to reduce all calculations (in spite of predistortions, noise in this channel remain maximum) will be

$$p_{nu} = \int_{F_k - \Delta F_k}^{F_k} kF^2\,dF \approx kF_k^2 \Delta F_k.$$

since $F_2 \gg \Delta F_k$.

The noise power in the group spectrum is /283

$$P_{ngr} = \int_{F_1}^{F_2} kF^2 dF = \frac{1}{3} k (F_2^3 - F_1^3)$$

or, using the preceding relationship for P_{nu}

$$P_{ngr} = \frac{P_{nu}}{3} \frac{(F_2^3 - F_1^3)}{F_2^2 \Delta F_k}.$$

Where $F_2 >> F_1$, as always occurs, we can consider $F_1 = 0$ without great error and then

$$P_{ngr} = \frac{P_{nu}}{3} \frac{F_2}{\Delta F_k}$$

or

$$\left(\frac{P_{ngr}}{P_{nu}}\right)_{(db)} = (F_2/\Delta F_k)_{(db)} - 5 \text{ db}.$$

The mean power of the group message P_{mgr} exceeds the power of the measured sinusoidal message in a channel by the following number of times:

$$\frac{P_{mgr}}{P_k} = 10^{b_{m(db)}},$$

and the peak power -- by the following number of times:

$$\frac{P_{peakgr}}{P_k} = 10^{b_{m(db)} + \nu_N}.$$

If we now reduce this result to a certain sine wave, equivalent in peak value to the group message, in order to use graph 9.3 it is necessary that

$$\frac{P_{eqgr}}{P_k} = 10^{b_{m(db)} + \nu_{N-3}},$$

or, considering standard predistortions (+4 db in the upper channel)

$$\frac{P_{eqgr}}{P_{kn}} = 10^{b_{m(db)} + \nu_{N-7}}$$

Thus, in order to determine the optimal modulation index, we should lay out the quantity

$$(P_s/P_n)_{eqgr} = 3(P_s/P_n)_k \frac{10^{b_{m(db)} + \nu_{N-7}} \Delta F_{gr}^2 \Delta F_k}{(F_2^3 - F_1^3)} =$$

$$= 3(P_s/P_n)_k \frac{10^{b_{m(db)} + \nu_{N-7}} \Delta F_k}{\Delta F_{gr}}$$

on the ordinate of Figure 9.3 and then continue as in § 9.1.

We assume in this case that the band of the FM line is approximately $2m_{gr}\Delta F_{gr}$, where the peak modulation index of the group message /284

$$m_{gr} = m_{effk}[20^{b_{0(db)}}],$$

where m_{effk} is the effective modulation index per telephone channel.

The assumption that the band is simply equal to twice the peak deviation can be considered by calculating nonlinear distortions as outlined in § 9.2; however, this consideration is generally unnecessary in selecting the optimal deviation.

Let us now study the most frequently used method of multistation operation with frequency division of signals from the different stations using frequency division multiplexing of channels and frequency modulation at each station.

The frequency deviation for each signal can be selected as is described in § 9.1, depending on the number of telephone channels transmitted by a given correspondent, the type of demodulator used and the fixed recommendations for thermal noise in the channel (this last figure should be replaced by the corresponding ratio of signal to noise in decibels in order to use the curves of § 9.1). Then, fixing the permissible additional crosstalk noise using formulas from § 9.2, we must select the transmission band for each correspondent on the basis of the noise from neighboring channels (see § 9.6) and the known IF amplifier selectivity -- the gap between signals and, finally, adding the bands of all signals and all gaps -- the total bandwidth of the HF trunk. Reduction of this band rapidly leads to an increase in relay station power, i.e., as in the case of transmission of the signal by only one station, there is an optimal HF band width. The problem arises of optimal distribution of noises resulting from the factors listed above. This problem is generally solved quite approximately, assigning an extremely small portion (15-20%) of the total permissible noise to all noise sources other than thermal. A portion of the permissible noise should also be allocated to noise over the Earth-satellite sector, interference from radio relay lines, etc.

However, as was shown in Chapter 8, when several FM signals pass through a common relay station, yet another type of noise arises -- crosstalk interference between signals of different correspondents. This noise results first of all from nonlinearity in the amplitude characteristic of the relay station, causing nonlinear products to appear in the bands of other signals. Usually, nonlinearity of the output device of the relay station -- the traveling wave tube -- is decisive.

Secondly, and practically to a still greater extent, crosstalk noise between correspondents results from conversion of amplitude modulation to phase modulation, characteristic for TWT.

The sequence of calculation of crosstalk noise produced in TWT is presented in Chapter 8. /285

The most important problem in selecting parameters of the HF line for the case of multichannel telephony with frequency division multiplexing is optimal distribution of recommendations between thermal noises and crosstalk noise of the latter type, resulting from multistation access. Solution of the problem depends not only on the overall throughput capacity of the system and the number of stations, but also on the specific characteristics

of the relay station (TWT); therefore, the solution can be produced only by trial and error, by successive calculation of a number of versions. Obviously, as the permissible thermal noise increases, it becomes possible to decrease signal power; thus, at the same time, we decrease crosstalk noise, since the TWT, operating in a more linear sector of its characteristic, creates less transient PM. If the total noise drops in this case, the number of channels transmitted can be somewhat increased, bringing the total noise to its required value (on the order of 7000 pw, considering allocation of a portion of the total requirement to the other noise sources mentioned at the beginning of the section).

If the resultant noise increases when the requirements for thermal noise are increased, the optimum must be sought in another direction.

Calculations performed for the TWT characteristics presented in [9.16] and for a total throughput capacity of the relay station of 600 telephone channels have shown that with five ground stations, approximately 400 pw is a reasonable allocation for thermal noise, i.e., somewhat more than half of all permissible noise. This division is typical for most cases. The optimum can be easily shifted to the area of higher noise if interference-stable FM signal detectors are used in the system; this greatly expands the required trunk band, but an increase in the number of ground stations does not cause rapid loss of total throughput capacity of the system.

There is also interest in the limiting case of a multistation system with frequency division of stations, when each telephone channel is transmitted on an individual carrier with FM, i.e., ordinary frequency division multiplexing of telephone channels is not used within the limits of one station, in order to give the system the greatest possible flexibility (low-channel system). In this case, the deviation of frequency for each carrier is generally selected on the basis of the noise level in a channel.

However, the next step -- calculation of the necessary HF band of each carrier -- is performed according to the recommendations of § 9.2 on the basis of the requirement for nonlinear distortions in the channel (Chapter 6), not the value of intermodulation noise. We recall that this requirement results from the use of a telephone channel in so-called secondary multiplexing systems. The distribution of noise in the channel between thermal and intermodulation noise, arising as the sum of the FM carriers passes through the TWT, /286 is performed on the same basis.

When time and multistation access are used, calculation of all parameters of the radio line is possible, based on the statements made above. For simplicity, we will assume that time access is organized by separate time multiplexing of signals of each telephone channel with sequential transmission during a portion of the

frame time (see Chapter 8), so that the redistribution of channels between stations within limits of the total number of channels does not change the form of the signal arriving at the input of the satellite relay station. We will also assume that multiplexing is achieved with no disruption of the analog structure of the signal, with factor Nq_t, where N is the number of channels, q_t is a coefficient considering the introduction of guard intervals between channels.

Then, at each moment in time the signal of a single channel is transmitted, and the interference stability of the system is determined by the method usual for FM, with the difference that the bandwidth of the channel is expanded by Nq_t times. The noise requirement considering this expansion determines the optimal index for modulation and frequency deviation (§ 9.1). Next, we must determine the bandwidth of the HF line (§ 9.4), based on the norm for nonlinear distortions in one channel, but considering the expansion of the band of frequencies of a channel. The band of frequencies of the HF line produced in this case may be found insufficient to eliminate intermodulation interference between channels resulting from linear distortions of the signal formed (formation of "tails," added to signals of other channels). This can be checked by calculating the linear distortions of the message transmitted by FM as is shown below in § 9.6. Particular difficulties arise in planning of combined systems with time multistation access, when time division multiplexing is performed for a signal which has been preliminarily frequency division multiplexed with a multichannel telephone message. Analysis of such systems was performed earlier in Chapter 8.

§ 9.4. Linear Characteristics of High Frequency Lines

During the first years of use of frequency modulation, it was falsely assumed that the non-uniformity of the frequency response of the HF section of an FM line would not produce low frequency distortions in the message since, it was said, the amplitude limiter eliminates all changes in signal amplitude which arise in the HF elements. The error in this naive conception is obvious -- the frequency detector reacts to the rule of change of signal frequency, not to its amplitude; the influence of the HF circuits on the rule of frequency modulation must be determined solving the problem of linear message distortions.

It has been shown [9.17] that in certain cases, if the asymmetry of the frequency responses of the circuit $K(\omega)$, $\phi(\omega)$ with positive and negative detuning is not too great, the modulus[1] and phase of the transmission factor for the modulating

[1]Having in mind the transmission factor normalized for its value at the resonant frequency.

process (with frequency Ω) are equal to:

$$K(\Omega)=\frac{1}{2}\{|K(\omega_0+\Omega)|+|K(\omega_0-\Omega)|\},$$

$$\varphi(\Omega)=\frac{1}{2}[\varphi(\omega_0+\Omega)-\varphi(\omega_0-\Omega)].$$

Thus, the modulus (phase) of the transmission characteristic at modulating frequency Ω is equal to the arithmetic mean of the values of modulus (phase) of the transmission characteristic of the circuit for high frequency with detuning at ±Ω relative to the carrier.

If the resonant curve of the circuit is symmetrical

then
$$K(\omega_0+\Omega)=K(\omega_0-\Omega),$$
$$K(\Omega)=|K(\omega_0+\Omega)|;\quad \varphi(\Omega)=|\varphi(\omega_0+\Omega)|,$$

i.e., the frequency response for low frequency simply repeats the form of half the frequency response of the HF circuit.

This conclusion can be explained by an illustrative discussion. Suppose the modulation index is much less than 1; we can then consider that the spectrum of the FM signal consists of the carrier (U_{car}) and one pair of side bands (U_s), separated from the carrier by the frequency of the modulating oscillation Ω. The resultant of the side bands is a vector perpendicular to the carrier (Figure 9.7).

The greatest phase deviation of the resultant oscillation

$$m = \phi_{max} = \arctan \frac{2U_s}{U_{car}},$$

while the greatest deviation of frequency is proportional to the phase deviation. Obviously, with changes in the side bands relative to the carrier by $|K(\omega_0 \pm \Omega)|$ times to the values U'_{su}, U'_{sl}, the deviation of phase and frequency will become less by the same number of times (dotted line on Figure 9.7). A similar conclusion

can be drawn directly from the vector diagram for phase $\phi(\Omega)$, and also in the case of the asymmetrical characteristic $K(\omega)$ [9.1]. Conclusions drawn with a low modulation index should remain correct as it is increased, so long as significant nonlinear distortions of the message do not arise, since in this case the transfer factor of the line remains practically constant.

In planning satellite communications lines, the modulation index is generally significantly greater than 1, while the transmission band of the HF line must be selected greater than $2\Delta\omega_d = 2m\Omega_u$ in order to retain the fixed nonlinear distortions. Therefore, within limits $(\omega_0 \pm \Omega_u)$, the frequency responses of the HF line are almost ideal, while the influence of the HF line on nonlinear distortions is not great. For this same reason, linear distortions in the frequency detector and modulator are slight, as long as we are concerned with the HF portion, whereas the portion of these devices through which the modulating signal passes can cause great distortions of the frequency response. /288

Things may be much more complex when interference-stable FM signal demodulators are used, since in order to achieve the greatest interference resistance and stability (in the sense of protection from self-excitation), special frequency feedback responses must be formulated (see [9.1]), due to which the equivalent frequency responses of such demodulators are often quite non-uniform. For example, in the case of a demodulator with FFB for reception of black and white television, the effective feedback is introduced in a band narrower than the band of frequencies of the message, since the spectral density of the message at high frequencies is slight and the decrease in depth of feedback does not result in nonlinear distortions, creating a gain in interference stability [9.18]. A decrease in feedback depth within the limits of the working band creates great non-uniformity of the resulting frequency responses of the device, generally requiring a special corrector.

In the case of transmission of multichannel telephony with multistation frequency access to the frequency responses of the overall line through which the multistation signal passes (i.e., to the on-board relay station and input portion of the receiver), no specific requirements are stated, since this line, transmitting the signals of all stations has very little unevenness of frequency responses within the limits of each band.

In this case, however, very rigid requirements are placed on non-uniformity of the amplitude-frequency responses of the transmitters of each station. Non-uniformity in these responses results in the formation of accompanying AM (due to the conversion of FM to AM), so that the AM is converted to PM in the traveling

wave tube, creating intelligible crosstalk interference between stations.

In order to retain this interference within the recommended magnitude, the parasitic AM arising due to non-uniformity of the amplitude-frequency responses of the individual lines generally should not exceed a fraction of 1% (calculation formulas presented in Chapter 8).

In the case of transmission of multichannel telephony with timed multistation access, the linear distortions also represent significant danger, since distortions of the leading edges of pulses arise, causing superposition of pulses of various stations on each other and, consequently, crosstalk interference. These distortions can be calculated on the basis of methods used for planning of multistation systems with time division multiplexing of channels. /289

Let us now study briefly the combination of properties of a receiving line determining its interference stability in relationship to interference from a neighboring channel. Detailed analysis of these properties goes beyond the scope of this volume, but the possibilities for construction of a satellite communications system depend to a significant extent on the capabilities of frequency separation of trunks from each other and the signals of individual stations within a trunk (with multistation frequency access). Therefore, estimation of the requirements for a line, determining the capabilities for frequency division, is a part of analysis of the problem of construction of the communications system. This is particularly true, since the requirement for good separation of FM signals on neighboring frequencies leads to the necessity of providing filters with high amplitude-frequency response cutoff steepness, which is harmful to both as to linear and particularly as to nonlinear distortions of the FM signal, greatly complicating the task of creation of phase correctors; this must be considered in determining the basic characteristics of line elements.

The problem of the effects of radio interference was solved in most general form in [9.19] for the case of frequency modulation of a signal and interference by multichannel telephone messages, when the spectra might partially overlap. Solution of the problem involves serious mathematical difficulties, due to which it was completed only for the case of the influence of ideal FM signals (without parasitic AM) and ideally linear FD with ideal amplitude limitation.

This case does not correspond fully to practical conditions, but is quite useful for estimation of the theoretical limit. Based on [9.20], where the solution for certain particular cases is continued to the point of numerical results, we can state that the interference which appeared was unintelligible, that the

signal/noise ratio at the output was proportional to the ratio of the carrier signal to noise at the input of the frequency detector, that the noise at the output drops rapidly with increasing detuning of the interfering signal relative to the useful signal. If we estimate the signal/noise ratio at the output with the most unfavorable detuning, i.e., the maximum noise, we can simply assume for a rough estimate that the ratio is equal to the ratio of the carrier signal and noise at the input of the FD. This maximum corresponds approximately to detuning near the boundary of the transmission band of the receiver for low frequency[1].

Actually, as long as the interference falls within the band of frequencies of the signal received, beating between carriers is most significant for the formation of noise at the output of the receiver. The variable portion of signal frequency under the influence of nonmodulated interference is determined by the relationship /290

$$\omega_{-}(t)=\frac{\delta\omega}{\frac{\rho-\cos(\delta\omega\cdot t)}{\frac{1}{\rho}-\cos(\delta\omega\cdot t)}+1}. \tag{9.7}$$

where $\delta\omega = \omega_2 - \omega_1$ is the difference between signal and noise frequencies, ρ is the ratio of their amplitudes. As we see, the interference at the output is proportional to the difference of frequencies and therefore increases with increasing detuning. However, if the beat frequency is higher than the pass band of the receiver for low frequency, the noise will be eliminated at the output of the receiver.

Analysis has shown that with frequency-modulated noise, the output of an ideal frequency detector carries no components with the modulation frequency of the noise and, generally, no components in the pass band of the receiver if the detuning is sufficiently great (spectra do not overlap).

However, interference still develops in an actual receiver. One reason for this is the non-ideal nature of the FD, particularly the nonlinearity of the detector characteristic.

With nonlinear detection, the voltage at the output includes a constant component. The value of this component is a function of the frequency and amplitude of the interfering signal. It

[1]More detailed quantitative results are presented in Chapter 10.

should be considered that the signal from the interfering station, operating at the frequencies of neighboring channels, passes through the receiver on the slope of the resonant characteristic, taking on practically 100% AM. The "constant" component now changes with the frequency modulation of the interference; intelligible and unintelligible transient interference arises, the value of which can be determined graphically or analytically based on the considerations discussed.

The second reason for the influence of radio interference outside the band is the nonideal nature of the operation of the amplitude limiter of the receiver, its limiter characteristic [9.21], due to which a constant component also arises in the output of the FD, dependent on the amplitude and frequency of the two signals.

The third reason for the development of interference is conversion of the AM to PM, occurring to some extent in any stage of the receiver, for example due to changes in the input capacitance of a tube or transistor. This process occurs most strongly in the limiter stages, operating in a clearly nonlinear mode. The resultant signal, modulated in amplitude with the frequency modulation of the interfering station, takes on angular modulation, which causes crosstalk interference.

Considering all of these factors, the filters of receivers must meet strict requirements; usually selectivity of IF filters /291
of 30-60 db at the carrier frequency of the nearest signal is sought, with small gaps between the bands of signals (about 20-40% of a band) and under particularly important conditions this selectivity is achieved at the nearest boundary of the spectrum of the neighboring signal. Some selectivity should also be provided by the input circuits of the receiver in order to avoid overloading of the input stages with strong interference from outside the band.

Input filters should also eliminate overloading of the input stages of a receiver by the signals of transmitters from the same ground station, particularly in those cases when the receivers and transmitters use a common antenna.

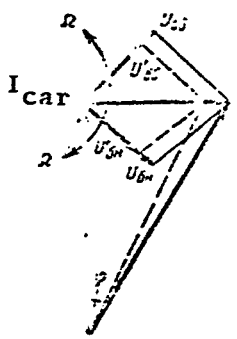


Figure 9.7. Vector Diagram with Low FM Index

REFERENCES

9.1. Kantor, L. Ya., Metody povysheniya pomekhozashchishchennosti priyema ChM signalov [Methods of Increasing the Interference Stability of FM Signal Reception], Svyaz' Press, 1967.

9.2. Malolepshiy, G. A., "Fluctuating Interference in Radio Lines with FM for Transmission of Multichannel Signals with Pulse-Phase Modulation," Radiotekhnika, Vol. XIII, No. 7, 1958.

9.3. Smirnov, V. A., Osnovy radiosvyazi na UKV [Principles of Radio Communications on USW], Svyaz'izdat Press, 1957.

9.4. Ageyev, D. V., "The Active Band of the Frequency Spectrum of a Function of Time," Trudy GPI im. Zhdanova, Vol. XI, No. 1.

9.5. Kantor, L. Ya., Yu. A. Afanas'yev, "Effectiveness of Limiting Position of Threshold and Maximum Deviation of Frequency in FM Radio Communications Systems," Elektrosvyaz', 1967, No. 10.

9.6. Zyuko, A. G., "Calculation of the Interference Stability of a Receiver with a High Level of Fluctuating Noise," Radiotekhnika, No. 10, 1956.

9.7. Shannon, C., "The Statistical Theory of Communications," Teoriya peredachi elektricheskikh signalov pri nalichii pomekh [Theory of Transmission of Electrical Signals in the Presence of Interference], Foreign Literature Press, 1953.

9.8. Levin, B. R., Teoriya sluchaynykh protsessov i yeye primeneniye v radiotekhnike [The Theory of Random Processes and its Applications to Electronic Engineering], Sovetskoye Radio Press, 1960.

9.9. Krivosheyev, M. I., Osnovy televizionnykh izmereniy [Principles of Television Measurement], Svyaz' Press, 1964.

9.10. Kart'yanu, G., Chastotnaya modulyatsiya [Frequency Modulation], Rumanian People's Republic Academy Press, 1961.

9.11. ICCR Recommendation No. 405.

9.12. Ryskin, E. Ya., "The Threshold Level of FM Receivers," Elektrosvyaz', 1964, No. 6.

9.13. Borodich, S. V., "Calculation of Noise in Radio Relay Line Channels with Frequency Division Multiplexing and Frequency Modulation," Elektrosvyaz', 1956, No. 1.

9.14. Polukhin, V. A., "Calculation of Noise of Nonlinear Transformations in Communications Lines with FM and Frequency Division Multiplexing for the Case of Phase Characteristic Correction," Sb. Trudov NIIR, No. 4(49), 1967.

9.15. Borodich, S. V., "Method of Calculation of Nonlinear Transitions in High Frequency Line of Multichannel Radio Relay Systems," Elektrosvyaz', 1967, No. 1.

9.16. Sunde, E. D., "Transient Distortion in FM Systems with Several Carriers," IEEE Convention Record, 1965, p. 13.

9.17. Borodich, S. V., "Linear Distortions in FM," Radiotekhnika, Vol. 9, 1954, No. 6.

9.18. Giger, A. I., I. G. Chaffee, "The FM Demodulator with Negative Feedback," BSTJ, July 1963, No. 4, p. 1.

9.19. Borodich, S. V., "Calculation of Permissible Radio Interference in Multichannel Radio Relay Systems," Elektrosvyaz', 1962, No. 1.
9.20. Medhurst, R. G., E. M. Nicks, W. Grasset, "Distortion in Frequency Division Multiplex FM Systems due to an Interfering Carrier," AIEE, Part B, May 1958.
9.21. Sidorov, V. M., "The Process of Transition to Reception of an Interfering Station in an FM Signal Receiver," Radiotekhnika, 1963, No. 11.

CHAPTER 10. MUTUAL INTERFERENCE BETWEEN SATELLITE COMMUNICATIONS SYSTEMS AND OTHER SYSTEMS

§ 10.1. Possible Cases of Interference and CCIR Recommendations for Permissible Interference /292

Chapter 5 presented a study of the basic factors influencing the selection of a band of frequencies for space communications, and also listed the bands of frequencies allocated for space communications internationally. It was pointed out in this same chapter that, since the bands of frequencies suitable for space communications services are broadly used at the present time by other services, the bands of frequencies for space services are allocated on the basis of joint utilization with those other services with which joint utilization is possible.

Preliminary studies of the capabilities of joint utilization have, as we know, frequently have led to determination of bands of frequencies in which services compatible with space communications services operate, to the development of limitations to be placed on both services, and to the establishment of recommendations for mutual interference. Joint utilization of frequencies has been possible basically with fixed services, namely line of sight radio relay lines. Thus, the study of mutual interference between line-of-sight radio relay lines and satellite communications services, as the most typical case, is of great practical importance both for calculation of specific cases of interference and for possible reanalysis of the existing recommendations and limitations.

It is the purpose of this chapter to analyze certain general considerations on the compatibility of various services and present a mathematical apparatus both for calculation of mutual interference and for critical evaluation of the temporary norms and limitations established by the CCIR.

The following four considerations were basically considered in developing these norms and limitations:

a) the satellite signal should not create impermissible interference in surface communications receivers;

b) the ground station signal should not create impermissible interference in surface communications receivers;

c) the signals from surface stations should not create /293
impermissible interference in satellite ground station receivers;

d) the signals of surface stations should not create impermissible interference in satellite receivers.

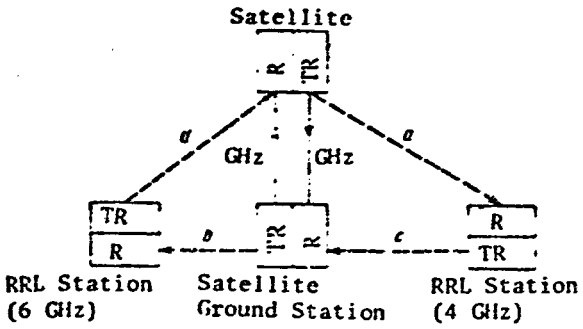


Figure 10.1. Possible Cases of Mutual Interference Between Satellite Communications System and Surface Radio Relay Systems

_____ useful signal, - - - - interference

Figure 10.1 shows possible cases of development of interference for satellite communications systems currently at used in the 4 and 6 GHz bands.

The problem of the possibility of combination of the two services, including radio relay systems and satellite communications systems, is determined by the following factors:

-- the maximum permissible interference in a telephone or television channel at the output of a system subject to interference;

-- the number of possible sources of interference, among which the permissible interference must be divided;

-- the ratio of powers or spectral power densities of useful and interfering signals at the input of the receiver for which the noise at the output of the receiver is equal to the permissible noise for the types and parameters of signal and noise modulation used;

-- the power or spectral density of power of the interfering signal;

-- the power or spectral power density of the useful signal;

-- the transmission losses over the propagation path of the interference (considering gain of the antenna, attenuation of the path and influence of polarization);

-- the losses upon transmission over the propagation path of the useful signal (considering antenna gain and attenuation over the path). /294

Recommendations will also be presented for permissible interference at the output of the system and the influence of other factors on the possibility of joint utilization of frequencies in the two services and the interference at the output in the specific case in question.

The recommendations for permissible interference in a telephone channel of a satellite communications system were established on the basis of the following considerations:

-- it is desirable that noise in the telephone channel of the satellite communications system resulting from interference by radio relay system transmitters represent a small fraction of the total noise in the system;

-- it must be considered that the satellite communications system may receive interference both through the satellite receiver and through the ground station receiver.

Considering these facts, the CCIR has recommended that satellite communications systems and radio relay systems operating in the same frequency bands be planned so that the power of interference at the point of zero relative level in any channel of the basic hypothetical standard circuit for the satellite system, caused by all transmitters of the radio relay stations be not over:

-- 1000 pw psophometrically weighted mean power for any hour;

-- 1000 pw psophometrically weighted mean power for one minute during more than 20% of any month;

-- 50,000 pw psophometrically weighted mean power for one minute during more than 0.03% of any month.

The recommendations for permissible interference in a telephone channel of the radio relay system were established similarly, but it was considered that the interference might arise either due to the transmitter in the satellite or due to a ground station transmitter.

Considering this, the CCIR has recommended that satellite communications systems and radio relay systems operating in the same frequency bands be planned so that the noise power resulting from the operation of ground stations and communications satellites at the zero relative level point of any channel of a hypothetical standard circuit of the radio relay system 2500 km in length not exceed the following values:

-- 1000 pw psophometrically weighted mean power for any hour;

-- 1000 pw psophometrically weighted mean power for one minute for over 20% of any month;

-- 50,000 pw psophometrically weighted mean power for one minute during more than 0.01% of any month.

The values presented relate to a telephone channel; the corresponding recommendation for permissible interference in a television channel has not been developed by CCIR. /295

Furthermore, there is still another important case of possible development of interference -- interference between two satellite communications systems operating at the same frequency. The problem is already arising of the mutual interference of such communications systems, and as their number grows the problem of joint

operation of several satellite communications systems may come to be one of the most important problems of planning.

Thus, the problem of mutual interference and joint operation of several different communications systems in the band of frequencies has not yet been fully solved. However, in spite of the fact that we present no such solution here, we should still analyze the method of solution of a more limited problem.

§ 10.2. Calculation of Noise Power in a Telephone Channel of a Multichannel Telephone Transmission System

Suppose P_{nk} (dbm 0) is the noise power in a telephone channel of a certain system, resulting from the influence of an interfering station of another system. Obviously, this value of P_{nk} should be related to some definite percentage of the time. For most receivers (with the exception of PCM receivers), power P_{nk} is proportional to the ratio of noise power to useful signal power at the input of the receiver P_n/P_s.

The proportionality factor in this case depends on the method of modulation of the noise and signal, the nature of transmission, loading conditions, difference in carrier frequencies, etc. This factor is called the receiver transfer characteristic, since it is usually fixed as a function of some parameter of the noise or signal and is represented by B_n.

Thus, at any moment in time

$$P_{nk} = \frac{P_n}{P_s} + B_n \text{ (db 0)}, \qquad (10.1)$$

where P_n/P_s and B_n are expressed in decibels.

Quantity P_n/P_s consists of the following components:

a) the power of the useful signal P_s at the output of the receiver antenna;

b) the power of the interfering transmitter P_i at the input of the transmitting antenna of the interfering station;

c) the sum of the gains of the receiving and transmitting antennas $(G_r + G_i)$ db over the propagation path of the noise; /296

d) attenuation over the path of the noise (L_0) between isotropic antennas.

Obviously,

$$P_n/P_s = P_i + G_r + G_i - L_0 - P_s \text{ (db)}, \qquad (10.2)$$

from which

$$P_{nk} = P_i + G_r + G_i - L_0 - P_s + B_n \text{ (dbm 0)}. \qquad (10.3)$$

Of all quantities in the right portion of the equation, P_i and P_s are usually constant, the gain of one of the antennas also remains unchanged with time, whereas the other parameters have a certain time distribution. The change in the remaining parameters with time is caused by the movement of the satellite (G), changes in propagation conditions (L_0) or changes in loading conditions (B_n).

Thus, simple addition in the form indicated in (10.3) is impossible. The change in certain parameters with time allows us to determine the distribution of P_n in time. To do this, addition of the variable parameters in (10.3) should be replaced by convolution of their probability densities, expressed in decibels.

§ 10.3. Influence of Propagation Conditions

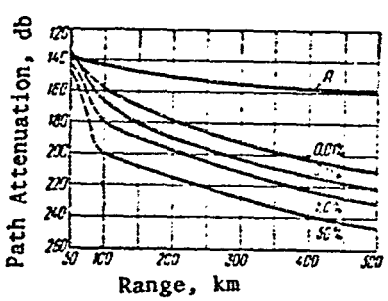


Figure 10.2. Attenuation of Path Between Isotropic Antennas not Exceeded for Percentages of Time Indicated over Curves (Surface Paths, Frequency 4 GHz). Curve A: Propagation in Free Space

Figures 10.2 and 10.3 show data relative to propagation which can be used to calculate noise over a surface path between a ground station and a radio relay station. Figure 10.2 shows attenuation curves for a path between isotropic antennas operating at 4 GHz. Figure 10.3 shows the correction factors for Figure 10.2 (for distances of over 100 km) allowing us to determine the transition to path attenuation for any frequency in the 1-10 GHz band. The information presented in Figures 10.2 and 10.3 was produced by measurements performed in moderate

and subtropical climates over paths with relatively even terrain with antennas at heights of 50 and 15 m.

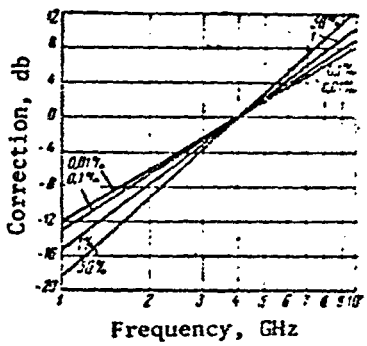


Figure 10.3. Correction for Curves of Figure 10.2 (Correction for Curve for 0.01% Suitable also for Curve A, Figure 10.2)

For distances of over 150 km and for other antenna heights, attenuation should be calculated from the point /297

$$l = 45 - 4.1(\sqrt{h_1} + \sqrt{h_2}), \text{ km}$$

where l is the true range;
h_1, h_2 are the antenna heights, m.

The portion of the curves relating to ranges on the order of 50-100 km was produced on the basis of a small file of data and the curves in this area should be used with great caution. This comment is also generally true of the remaining portions of the curves, since they were produced under average conditions and may show rather significant deviations in practice.

§ 10.4. Gain of Ground Station Antenna

In the general case, when moving satellites are used in a satellite communications system, the gain of the antenna in the direction of the radio relay station creating or experiencing interference will change with time.

The angle between the direction toward the radio relay station and the axis of the main lobe of the ground station antenna can be determined from the formula

$$\phi = \arccos [\sin \beta_s \sin \beta_r + \cos \beta_s \cos \beta_r \cos (A_s - A_r)], \quad (10.4)$$

where β_s is the angle of elevation of the satellite relative to the ground stations;
β_r is the angle of elevation at which the radio relay station is visible;
A_s is the azimuth of the satellite relative to the ground station;

A_r is the azimuth of the radio relay station relative to the ground station.

With known parameters of movement of the satellite, this same formula can be used to determine the time distribution of the angular difference between the direction of the antenna toward the satellite and toward the radio relay station. If the radiation pattern of the ground station antenna is also known, the distribution of gain of the ground station in the direction of the radio relay station can also be determined. If the radiation pattern of the ground station antenna is unknown, the standard radiation pattern for a ground station antenna, developed specially for purposes of study of mutual interference and recommended by ICCR, can be used (Figure 10.4).

According to the suggested pattern, the gain of the antenna is independent of frequency and antenna size within the limits of the main lobe and is described by the approximate expression

$$G_s = 32 - 25 \log_{10} \varphi, \text{ db.} \qquad (10.5)$$

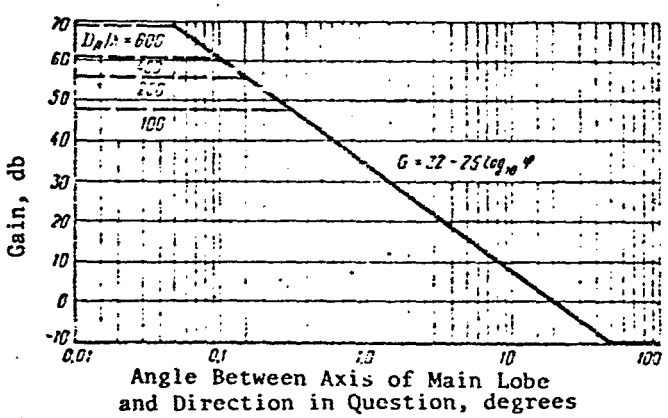


Figure 10.4. Suggested Standard Ground Station Radiation Pattern (D_A/λ is the ratio of Antenna Diameter to Wavelength)

§ 10.5. Transfer Characteristics of Receiver

Calculation of the transfer characteristics of a receiver is a rather difficult task, since even with assigned methods of modulation, a large number of parameters of the interfering and useful signals must be involved.

Efforts were primarily directed toward calculation of the transfer characteristics for transmission of multichannel telephony in systems with frequency and phase modulation. This is explained by the fact that frequency modulation is broadly used in radio relay lines and all experimental satellite communications systems also use frequency modulation.

The many studies in this area [10.1-10.6] have suggested many methods of calculation of various degrees of complexity and under various assumptions. Let us discuss one of them, suggested by S. V. Borodich [10.6], the most complete and detailed.

Suppose the input of an FM receiver is acted upon by a useful signal and noise:

$$u_s(t) = U_s \cos [\omega_s t + \alpha_s(t)]$$

$$u_n(t) = U_n \cos [\omega_n t + \alpha_n(t)], \qquad (10.6)$$

where U_s, U_n are the amplitudes of the useful and noise signals at the input of the receiver;

ω_s, ω_n are the carrier frequencies of the useful and noise signals; /299

α_s, α_n are the instantaneous phase increments of the useful signal and noise.

If the useful and interfering signals are frequency modulated, U_s and U_n are constant and

$$\alpha(t) = \int_0^t ku(t)\,dt, \qquad (10.7)$$

where ku(t) is the instantaneous frequency increment;
k is the steepness of the modulator characteristic;
u(t) is the modulating signal.

If the amplitude of the noise voltage is less than the amplitude of the signal voltage and the noise is in the transmission band of the receiver, then

$$U_n/U_s = K < 1, \quad \omega_n = \omega_s + \delta\omega, \qquad (10.8)$$

(where $\delta\omega$ is the difference between noise and signal frequencies) and the resultant oscillation can be written as

$$U_s \cos(\omega_s t + \alpha_s) + U_n \cos(\omega_n t + \alpha_n) = U_\Sigma \cos(\omega_s t + \alpha_s + \theta), \tag{10.9}$$

$U_\Sigma = U_c\sqrt{1 + K^2 + 2K\cos\varphi}$ is the resultant amplitude of the voltage;

$\theta = \operatorname{arctg}\left(\frac{K\sin\varphi}{1 + K\cos\varphi}\right)$ is the parasitic phase increment caused by the noise;

$\phi = \alpha_n - \alpha_s + \delta\omega t$ is the instantaneous phase difference between noise and signal.

If the pass band of the receiver is wide enough that the spectra of the signal and noise are not distorted, if no significant combination oscillation arising in the nonlinear devices of the receiver fall within the pass band of the IF amplifier and, finally, if the limiter and discriminator of the receiver are ideal, the output of the receiver will carry the signal

$$u_{вых}(t) = u_s(t) + \frac{1}{k_s}\frac{d\theta}{dt} = u_s(t) + \varepsilon(t), \tag{10.10}$$

where the steepness of the demodulator characteristic is taken as the inverse of the steepness of the modulator.

Thus, determination of the noise at the output is reduced to determination of the statistical parameters of $d\theta/dt$ (or θ). In the case of transmission of a multichannel message, we are interested in the energy spectrum $\varepsilon(t)$, from which we can easily find the mean value of noise power in a telephone channel. In the case of transmission of one-channel messages, it is usually sufficient to know the mean value of noise power in the band of the useful message, i.e., the integral of the energy spectrum $\varepsilon(t)$.

Where $K < 1$, the parasitic phase increment of the useful signal, caused by noise, is not over $\pi/2$, i.e., $-(\pi/2) < \theta < +(\pi/2)$, and is a periodic function of ϕ, i.e., the instantaneous phase difference between signal and noise. /300

Consequently, θ can be represented as

$$\theta = \sum_{n=1}^{\infty} (-1)^{n-1}\frac{K^n}{n}\sin(n\varphi). \tag{10.11}$$

Let us find the correlation function of θ by averaging over the set

$$\overline{\theta\theta_\tau} = \overline{\sum_{n=1}^{\infty}(-1)^{n-1}\frac{k^n}{n}\sin(n\varphi)\sum_{m=1}^{\infty}(-1)^{m-1}\frac{k^m}{m}\sin(m\varphi_\tau)}. \tag{10.12}$$

The general term of the sum beneath the line for averaging over the set contains

$$\sin(n\varphi)\sin(m\varphi_\tau) = \int\int\int\int_{-\infty}^{+\infty}\sin(n\varphi)\sin(m\varphi_\tau)\,p\ (\alpha_s\alpha_{s\tau})\ \times$$

$$\times\ p\ (\alpha_n\alpha_{n\tau})\,d\alpha_s\,d\alpha_{s\tau}\,d\alpha_n\,d\alpha_{n\tau}. \tag{10.13}$$

In this last equation, we assume, as usual, that the mutlichannel message is a normal stable random process, processes $u_s(t)$ and $u_n(t)$ being independent.

Then, random processes $\alpha_n(t)$ and $\alpha_s(t)$ are also independent and normal. In this case, the two dimensional probability distribution density of the random quantity α is

$$p(\alpha,\ \alpha_\tau) = \frac{1}{2\pi\psi_{\alpha 0}\sqrt{1-R_\alpha^2}}\left(\frac{\alpha^2-\alpha_\tau^2-2R_\alpha\alpha\alpha_\tau}{2\psi_{\alpha 0}(1-R_\alpha^2)}\right). \tag{10.14}$$

where $R_\alpha = \psi\alpha(\tau)/\psi\alpha(0)$ is the correlation factor of the random processes α and α_τ;

$\psi\alpha(\tau)$ is the correlation function of α;
$\psi\alpha(0)$ is the dispersion of α.

As a result of integration and time averaging, we produce

$$\sin(n\varphi)\sin(n\varphi_\tau) = \frac{1}{2}\cos n\delta\omega\tau\exp\{-n^2[\psi_S(0)-\psi s(\tau)+$$
$$+\psi n(0)-\psi n(\tau)]\}. \tag{10.15}$$

Thus,

$$\overline{\overline{\theta\theta}}_\tau = \frac{1}{2}\sum_{n=1}^{\infty}\frac{K^{2n}}{n^2}\exp\{-n^2[\psi_S(0)-\psi_{aS'}(\tau)]- \\ -n^2[\psi_\Pi(0)-\psi_n(\tau)]\}\cos n\delta\omega\tau. \tag{10.16}$$

The averaged energy spectra of θ in this case will be /301

$$\overline{W}_\theta = 4\int_0^\infty \overline{\overline{\theta\theta}}_\tau \cos\Omega\tau\, d\tau \tag{10.17}$$

and according to the theorem of the spectrum of a derivative, we have

$$\overline{W}_\varepsilon(\Omega) = \frac{4\Omega^2}{k_S^2}\int_0^\infty \overline{\overline{\theta\theta}}_\tau \cos\Omega\tau\, d\tau.$$

Going over to the noise power in a telephone channel with mean frequency Ω_k and considering psophometric weighing [10.6], we produce

$$P_{nk} = \frac{\Delta F_k k_p^2 \Omega_k \cdot 10^9}{\Omega_2 \cdot \Delta\omega_{dkeff}^2}\sum_{n=1}^{\infty}\frac{K^{2n}}{n^2}[g(nq+b)+g(nq-b)], \quad nw, \tag{10.18}$$

where ΔF_k is the effective band width of a telephone channel;

k_p is the psophometric coefficient;

$\Delta\omega_{dkeff}$ is the effective frequency deviation of one channel;

$q = \delta\omega/\Omega_2$, $b = \Omega_k/\Omega_2$;

Ω_2 is the upper frequency of the group band.

It is assumed here that $\Omega_{2s} = \Omega_{2n}$.

In the brackets, the summation sign is followed by components such as $g(nq \pm b)$, which also depend on the modulation indices of the signal and noise:

$$a = n^2(m_c^2 + m_p^2),$$

where

$$m = \Delta\omega^2_{dkeff} e^{2b_m}/\Omega_2^2.$$

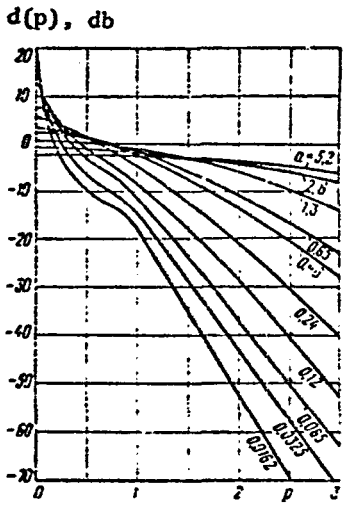


Figure 10.5. Calculation of Noise Power

The function $g_n(a, nq \pm b)$ is shown graphically in Figure 10.5, where the parameter p replaces $nq \pm b$.

As follows from Figure 10.5, this function is rather complex. For large a and $0 < P \leq 10$, we can use the approximate formula

$$g(a, p) \approx \sqrt[1]{\frac{\pi}{2a} e^{-\frac{p^2}{2a}}}$$

Formula (10.18) yields the full solution of the problem. Calculations show that the maximum of interference in each channel is reached with detuning equal to the middle frequency of the channel in question, and also increases as the upper boundary of the group band is approached. /302

Thus, the maximum noise will be observed in the highest frequency channel with $\delta\omega = \Omega_2$.

Formula (10.18) is rather cumbersome, but can be somewhat simplified and it can be considered with accuracy sufficient for approximate calculations that for the highest frequency channel

$$P_{nkmax} = (3\Delta F_k k_p^2 \cdot 10^9/\Delta\omega^2_{dkeff}) g_1(0) K^2, \text{ pw}$$

or since $K^2 = P_n/P_s$in (see § 10.2)

$$B_{nmax} = (3\Delta F_k k_p^2 10^9/\Delta\omega^2_{dkeff}) g_1(0).$$

REFERENCES

10.1. Medhurst, R. G., E. M. Hicks, W. Grosset, "Distortion in Frequency Division Multiplex FM Systems due to an Interfering Carrier," AIEE, Vol. 105, Part B, May 1958, p. 282.

10.2. Chamberlain, J. K., R. G., "Mutual Interference Between Communication Satellite and Terrestrial Lin-of-Sight Radio-Relay System," IEE International Conference on Satellite Communications, Nov. 1962, Conference Digest, pp. 39-45.

10.3. Hamer, R., "Radio Interference in Multichannel Telephony FM Radio Systems," AIEE, Vol. 108, Part 3, Jan. 1961, p. 75.

10.4. Curtis, H. R., "Interference Between Satellite Communication Systems and Common Carrier Surface Systems," BSTI, Vol. 41, p. 921, 1962.

10.5. Shinji Hayashi, "On the Interference Characteristics of the Phase Modulation Receiver for Multiplex Transmission," Japan J. IECE, Vol. 35, pp. 522-528, Nov. 1952.

10.6. Borodich, S. V., "Calculation of the Permissible Value of Radio Noise in Multichannel Radio Relay Systems," Elektrosvyaz', 1962, No. 1.

CHAPTER 11. PRINCIPLES OF CONSTRUCTION OF CONTROL APPARATUS IN SATELLITE COMMUNICATIONS NETWORKS

§ 11.1. Purpose and Functions of Control Systems /303

Earlier, we studied the principles of design of communications networks using satellites. It was shown that satellite communications networks have a number of peculiarities, distinguishing them from surface networks and, which is particularly important, requiring the construction of such networks in the form of individual, autonomous systems having independent structure and organization. At the same time, main line satellite communications networks form a portion of the total communications network, performing definite tasks in this network and, consequently, are closely interrelated with the total network.

During the process of operation of a satellite communications network, orders for transmission of information from subscribers of the overall communications network, for organization of the required communications channels, for connection of these channels with the channels of the overall network, etc. are continually received. The network is in a continual process of change, supported by its control system. It should be noted that the control must be such as to assure optimal conditions for the transmission of information and maximum effectiveness of the operation of the network itself.

Thus, the purpose of the control system of a satellite communications network is to control the operation of the channels of the network so that the maximum possible volume of information is transmitted at any time with the required servicing quality and reliability. To perform this task, the control system must control the operation of the communications channels within the network and cause the satellite network to interact with the overall communications network. The tasks to be performed by the control system during the operating process are determined to a significant extent by the type of information transmitted. From this standpoint, we can arbitrarily divide the information transmitted in a satellite communications network into three main types:

1) general or broadcast-type information; /304

2) information transmitted over channels provided in advance;

3) information, the channels for which are provided immediately before transmission.

Let us study the functions which the control system must perform in the process of transmission of each of these types of information in more detail.

General information is generally transmitted to the entire network or a portion of the network. Transmission is most frequently from the central station of the network, but this information may be transmitted from peripheral stations as well. The time of transmission is usually known in advance from a schedule, although the necessity for unscheduled general transmissions sometimes arises.

Thus, the transmission of general information involves the operation of many stations of the network, and control of general communications is therefore best performed at the control center (CC) of the network. Control operations consists basically of the following:

a) reception of orders for the transmission of general information;

b) recording of orders received in memory (usually long-term memory) at the CC;

c) input of data on the necessary control of general information to the CC scheduling equipment;

d) output of instructions from the CC to the switching equipment at the station (central or peripheral) for reception of incoming general information from surface lines and switching of the information to the transmitting devices of the station;

e) output of instructions from the CC to the switching equipment of the stations of the network where the general information is to be received for connection of the proper receiving apparatus to the surface lines originating at the station;

f) output of instructions to switching equipment at the stations for return of apparatus to its initial condition after completion of the general transmission.

When channels prepared in advance are used, information, generally duplex, is transmitted according to a schedule known in advance or a preliminary order. However, this class also includes cases when the necessity for transmission of information arises unexpectedly, for example in case of a failure of surface lines, but in this case the request for provision of communications lines arrives at the CC of the network and the time available to arrange the necessary channels is not too rigidly limited -- at least some tens of seconds are available.

It is characteristic for this type of communications that channels are assembled on the scale of the communications network as a whole, since requests for channels are sent to the control center and fulfilled by the control center considering the /305
requirement for maximum effectiveness of operation of the entire network.

We should add to the above that, as a rule, we are speaking not of individual telephone channels, but rather of bundles or groups of channels used for transmission of large volumes of information at high speeds.

The basic operations performed by the CC in the provision of such channels consist of the following:

a) reception of a request for channels and its recording in memory and the scheduling equipment;

b) analysis of the situation in the network and determination of the possibility of provision of (duplex) channels at the required stations of the network. A version of organization is selected for which the effectiveness of operation of the network is maximal;

c) output of instructions to the switching equipment at all stations participating in operation to make the necessary connections for transmission and reception;

d) output of instructions to break connections after the subscriber requests that communications be interrupted.

The operation of a network of automatic switched channels, i.e., channels organized immediately after reception of a request for communications, is most complex in organization and control. In this case, connections must be established in accordance with each call arriving at each station. The station may operate simultaneously with many correspondents, and a call arriving may be directed to any one of them. Furthermore, in establishment of a connection, the type of channel required and category of call must be considered and, naturally, the channel must be organized so as to assure most effective operation of the station and of the network as a whole.

The number of connections established in the network over a comparatively short time interval may be very great, creating additional complication for operation in this mode. The large volume of operations required to analyze the situation and make connections in many cases makes creation of a centralized control system impossible. It will be necessary to develop decentralized systems, in which each station need not know the situation in the entire network, producing the information required in the process of making connections. In multichannel satellite networks, in addition to the provision of individual channels, it is also necessary to control subtrunks, i.e., groups of channels. This function has a particularly strong influence on the effectiveness of operation of a network and is therefore quite important. Groups are not controlled in multichannel networks as frequently as individual channels and, considering the importance of this operation, it should always be performed considering the entire situation in the network, i.e., by the centralized control system.

Thus, in a communications network using multichannel groups and operating in the mode of automatic establishment of connections upon subscribers requests, yet another problem arises -- matching of the operation of the centralized control system for groups of channels and the decentralized control system for individual channels. /306

The primary operations in the process of control of a network with automatically switched channels can be reduced to the following:

a) reception of requests for establishment of connections (a request for service arriving at the station of the satellite network from a surface communications center with the number of the subscriber being called);

b) analysis of the number and determination of the direction in which a channel should be organized;

c) search for a free channel in this direction;

d) sending of a call and the necessary information to the station with which communications should be established;

e) reception of the call at the called station and organization of a return channel;

f) checking of the established connection and switching of the channel to the subscribers (first stage of this connection consists in transmitting the number of the subscriber called from the calling station to the called station over the channel organized);

g) testing of the loads on all stations of the network;

h) in multichannel networks, sending of requests for additional subtrunks (groups) from ground stations to the control center in those cases when the arriving load cannot be serviced by the available capacity;

i) at the control center of the network, analysis of the situation in the network and making of decisions concerning transfer of additional subtrunks to various stations; the decision is made considering the provision of the maximum effectiveness for the operation of the network as a whole;

j) sending of instructions from the CC to ground stations concerning necessary connections in subtrunk switches and reception of reports of performance of instructions.

We present below a description of the functional systems and operating algorithms of devices performing the control functions outlined above in satellite communications networks.

It should be noted that all of the systems described below can be made as ordinary control systems for automatic telephone networks based on the use of relay or semiconductor equipment.

In addition to the functional systems, the remaining parts of the apparatus differ little from that used for automatic control in telephony, automation or computer engineering. In view of this and in order to reduce the length of this book, no more /307

detailed description of the control apparatus will be presented here.

§ 11.2. Control in Broadcast-Type Communications Networks and Communications Networks Using Pre-Assigned Channels

The principal control devices in broadcast communications networks are located at the control center of the network. Figure 11.1 shows an approximate block diagram of such a center.

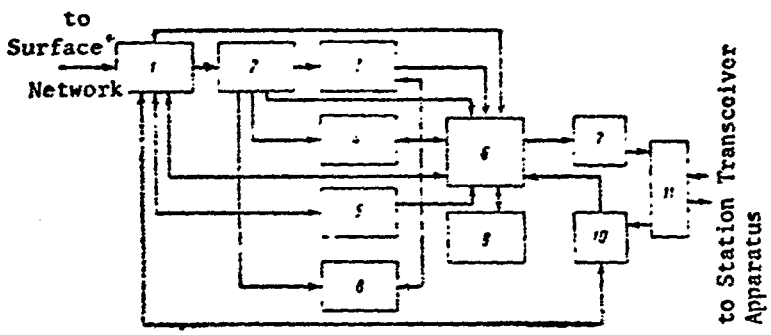


Figure 11.1. Control Center of Broadcast Satellite Communications Network: 1, Request Receiving Device; 2, Device Decoding and Analyzing Requests; 3, Device Recording Requests; 4, Main Memory Unit; 5, Device Recording Situation (Load) on Network; 6, Control Device; 7, Device Issuing Instructions to Stations of the Network; 8, Scheduling Device; 9, Network Load Optimizing Device; 10, Report Receiving Device; 11, Service Communications Transceiver

Requests for a channel arrive at the CC primarily over surface lines connecting the CC to the control centers of the surface communications networks which utilize satellite communications channels. Sometimes, calls may arrive from ground stations of the satellite network over the service communications lines in this network.

The calls arriving at the CC are sent to input device 1, where they are recoded. The reason is that calls arrive at the CC in the form in which they are transmitted in surface communications networks. This may be either a verbal message by telephone, or a message transmitted in the form of a telegram, in code, etc. In order to allow the message to be processed in the CC, it must be converted to some standard code used in the control devices of the CC.

After recoding, the call is sent to analyzer 2. Here the content of the request is determined, i.e., the type of channel, the information to be transmitted, the urgency and the category of the information transmitted, transmission time, point (or points) of origination of transmission and points (stations) of reception, etc. The results of the analysis are recorded into memory unit 3, which should be a long-term memory unit. However, there must also be a certain volume of operative memory 4 for processing of urgent requests, as well as to support the operation of logic device 6. /308

Quite frequently, the transmission of broadcast-type information is systematic, according to a schedule composed in advance and rarely changed. This type of transmission is used for television and radio broadcasts, newspapers, etc. Operation according to the schedule requires that the CC include a special device 8, which puts out instructions for transmission of various types of information at certain moments in time -- date, hour, minute, second. The operating program of this device is recorded in its long-term memory for each request.

In order to allow the control (logic) device of the CC 6 to make decisions concerning the provision of various channels, it must have, in addition to the request, information on the status of the network, i.e., the status of each channel of the network (busy, free, out of order, etc.), as well as the status of the transmitting and receiving equipment of the station (including terminal devices) and the possibility of connection from the ground station to surface communications networks. All of this information must be supplied by the ground stations to the CC over the service communications lines and recorded in network situation long-term memory 5. To avoid excessive loading of the service channels, this information should be transmitted only when changes of status occur at the stations of the network.

The control device (CD) analyzes the situation and makes a decision concerning organization of a channel for transmission of some type of information after arrival of a start signal from the scheduling device or, if the request is urgent, from the analyzing device. In order to make a decision, the CD selects a suitable channel by inspecting the situation concerning loading of channels and their suitability for operations in memory. During the process of search for a free channel, the operation of the network is also optimized using the load optimizing device 9. For example, if one of several free channels could be used, the channel is selected which is optimal as concerns power, frequency band and other parameters, while those which could be effectively used for transmission of other more complex types of messages are left in reserve. A channel is also selected which can be received at all stations of the network to which the information is to be send, since in a large network the ground stations as a rule cannot receive all possible channels used in a network, but rather

have equipment designed for reception (and transmission) of only a portion of these channels according to the traffic and to their actual capabilities. /309

The decision made in the CD consists of instructions to make the necessary connections to the switching devices of the stations of the network. Of course, different instructions are sent to the station which is to transmit and to the stations which are to receive. All instructions from the CD are sent to the instruction output device, where they are encoded as the corresponding code signal and supplied with addresses. The signals are then sent to the service communications transmitter 11.

It was noted above that the CC receives information on the status of all stations of the network. In addition to this, the CC receives reports of performance of instructions sent out, requests for channels from ground stations and other information. All of this information is sent to the service communications receiver, then decoded in receiver 10 and sent out according to the purpose of the signal received either to the CD or to the situation memory, the request input device, etc.

The main part of the control system at ground stations is the commutator. A signal arriving over a surface line at a ground station must be transmitted over one of the channels of the station. This means that the surface line must be connected to the proper transmitter of the station. A signal received from the satellite must be transmitted into a surface communications line, requiring that the output of a receiver be connected to the input of the proper outgoing surface line. The station commutator, both the receiving and transmitting portions, consists of devices with N inputs and R outputs, and any input can be connected to any output or group of outputs. The commutator is controlled using a special logic marker circuit, to which the decoded commands from the CC are sent.

The design of the commutator may be quite varied. The commutation elements may be relays of various types, hercons, or contactless elements. The primary requirement for a commutator, in addition to reliability of connection, is the provision of the necessary channel separation, which is achieved by proper selection of a circuit, switching elements, commutator control method and by its design.

In addition to the commutator, the control system of a station of the network includes devices receiving instructions from the CC and decoding them, as well as devices which test connections made then transmit reports of connections made over the service communications line.

An approximate block diagram of the control apparatus is shown in Figure 11.2. As we can see from the figure, instructions arriving from the CC are fed from service communications receiver 1 to device 3, where they are decoded (at the same time their address is determined). The instructions are then sent to logic actuating device and marker 4, which determines the best version of connection in receiving commutator 6 or transmitting commutator 7 (or both). In some cases, some information from the CC may have to be recorded; memory device 5 is used for this purpose. /310

Testing of connections made with commutators is the function of test device 9, which transmits the results in the form of reports to device 10, in which the reports are encoded and supplied with addresses. The signals are then transmitted to service communications transmitter 2.

Requests for channels and other service information arrives from the surface network serviced by a given ground station are received by special device 8. If necessary, this information is sent to the memory device. If an incoming request is intended for the CC, it is encoded and transmitted to the CC through the service communications channels.

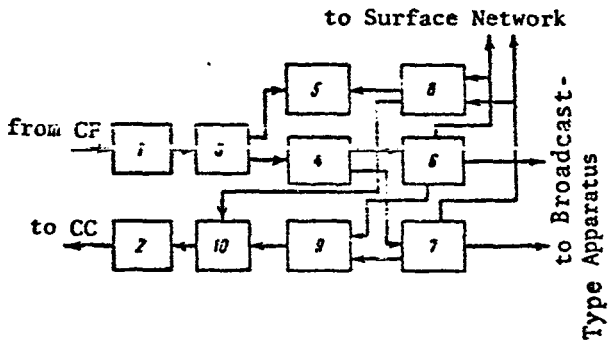


Figure 11.2. Ground Station of Circular Satellite Communications Network: 1, Service Communications Receiver; 2, Service Communications Transmitter; 3, Instruction Encoding Device; 4, Marker; 5, Long-Term Memory Unit; 6, Receiving Commutator; 7, Transmitting Commutator; 8, Device Receiving Service Information from Surface Network; 9, Connection Testing Device; 10, Device for Coding and Transmission of Instructions through Service Communications Line

The operation of a network with communications channels assigned in advance by schedule or preliminary request is quite similar to the operation of a broadcast-type communications network.

The primary distinction of this network from a broadcast network is that duplex radial channels are provided rather than simplex channels, generally connecting some pair of stations. Furthermore, the variety of types of channels used in the network is significantly greater than in a network for transmission of broadcast information. The CC of the control system in a network /311
with channels assigned in advance provides a duplex channel, for which connections are made simultaneously at each pair of stations in both the receiving and transmitting commutators. This operation requires more complex analysis of the situation in the network, since the CC must find two free simplex channels simultaneously and pair them into a duplex channel by simultaneous switching at two ground stations. Therefore, the volume of the memory device at the CC and ground station in the control system of this network will be greater than in a broadcast network. The control device of the CC and the markers at the ground stations must be somewhat more complex.

However, the general block diagrams of the control apparatus and algorithms of its operation will not differ practically from those described earlier in Figures 11.1 and 11.2. Therefore, we judge, more detailed description of these devices need not be presented here.

As we see, the control of networks for transmission of broadcast information and networks operating with channels assigned in advance is comparatively simple. The apparatus used for this control follows simple algorithms and need not be highly complex. It is significantly more difficult to solve the problems of control in networks using automatically commutated channels, i.e., in networks where a communications channel is not organized in advance, but rather after arrival of a call. In this case, depending on the method of organization of the satellite network, various versions of the control system may be used; as a rule, they are all more complex and cumbersome than those described above. We present below a description of some of the main versions of such systems.

§ 11.3. Control System in Low Channel Network with Centralized Control

A general diagram of the operation of the network is presented in Figure 11.3. Each ground station (GS) turns directly to the control center of the network to obtain a channel when a call arrives from the surface communications lines. The CC determines the possibility of providing a circuit, finds a free channel suitable for servicing of the incoming call and performs commutation simultaneously at the calling ground station GS-1 and the called ground station GS-2, after which the circuit formed is turned over to the subscribers for use.

A block diagram of the control center of the network is shown in Figure 11.4. The information from the GS received over the service communications lines (requests, reports, test information, etc.) moves from the service communications receiver 2 to the decoding device 4, where the address of each message and the address of the GS which has sent it, type of message, etc. are determined. In this same device, the signal received is decoded and sent to the proper section of the CC. Requests for channels, directly or through a supplementary memory unit 5, are sent to control device 10, i.e., to the logic device which develops the decision of the CC. /312

To make its decision, the CD turns to long-term memory unit 6 and operative memory unit 7, which contain information on the status of all channels in the network. Furthermore, the CD uses information reflecting the status of the quality of operations of channels of the network 9 and information recorded in the scheduling device 8.

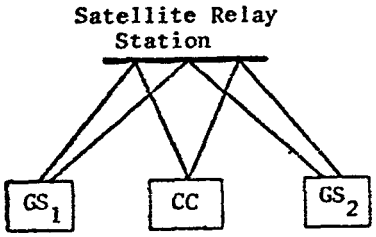


Figure 11.3. Operating Diagram of Low-Channel Network with Centralized Control

After analyzing the situation, the CD makes a decision on provision of a channel or sending of the busy signal to the subscriber. The decision is sent to the instruction-producing device 11, where it is supplied with an address, then to coding unit 13 and further to service communications transmitter 3, which transmits the instructions to the GS.

In order to increase the operating reliability of the service communications system, the instructions transmitted may be tested. To do this, a GS, after receiving an instruction from the CC, transmits it back to the CC, where a special test device compares the instruction received back from the GS with that transmitted. If the instruction received back from the GS corresponds to the instruction transmitted, the test device sends out an actuate command to the GS, after reception of which the primary instruction of the CC is performed. If the test instruction does not correspond to the initial instruction, the initial instruction is repeated. The return service communications line from the GS to the CC 14 is also tested.

The CC can receive requests for channels not only from GS, but also from the surface network 1. In this case, the operation of the CC is quite similar to that described above. The CD, in addition to instructions to the ground stations, also generates

and transmits response messages concerning the provision of a channel, busy signals, etc. into the surface network.

In the process of provision of a channel, the CC first sends an instruction to make the required connections to the GS, to the commutators of the called and calling stations. After these instructions are sent to the GS, the stations send reports to the CC. After receiving the reports and being sure that the connections are correct, the CC sends an instruction to the calling station to transmit the number of the called subscriber to the called station, which signals that the channel is ready for operation.

In the operation of a satellite communications network operating together with a surface network using automatic switching /313
of channels, it is very important for the "busy" signal from subscribers to be transmitted rapidly from the GS to the surface network. This is necessary to prevent disruption of the operation of the surface network control equipment, which usually determines whether channels are busy very rapidly.

However, in a satellite network with the operating algorithms described above, the CC determines whether a channel is busy; therefore, in order to learn whether it can provide a channel according to the request, it must send a request to the CC and receive a response. The time necessary for this is significantly greater than the time necessary to determine whether a channel is busy in surface networks.

It was shown above that in a network with centralized control, failure to provide a channel occurs when there are no free channels on the satellite or there is no free channel apparatus at the calling or called GS. One possible method of rapid production of a busy signal at the GS is that the CC continually informs each GS of the availability of free channels to the satellite and /314
all remaining GS in the network. In this case, the calling GS need only have information as to whether there are free channels to the satellite and to the called GS. Specifically which channels are available, what sort of switching must be performed to occupy them and all other necessary information for establishment of a connection can be produced at the GS over a longer period of time during the process of organization of a channel. If the GS has the information that the satellite or called station has no free channel, this is sufficient to give a quick busy signal to the subscriber.

Information on the availability of free channels is comparatively slight in volume and can be easily reported by the control center to all GS. In order to transmit this information, the CC should have special device 12, which will send messages to all GS

only if all channels are busy at the satellite or some ground station. Information of this type is recorded in the memory of all GS and considered in processing calls arriving from the surface network.

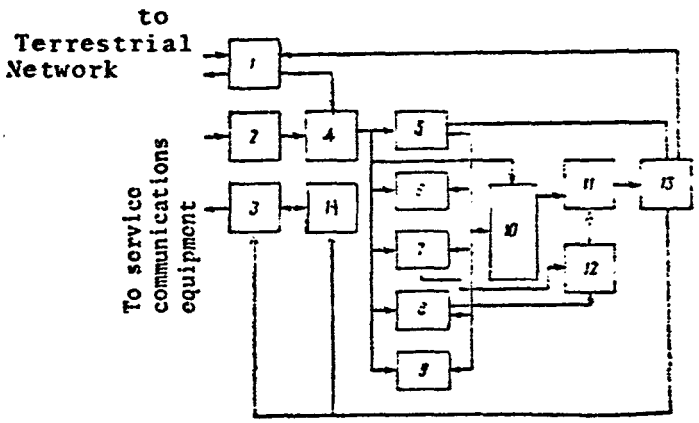


Figure 11.4. Low Channel Satellite Communications Network Control Center: 1, Device for Service Communications with Surface Network; 2, Service Communications Receiver; 3, Service Communications Transmitter; 4, Device for Decoding Service Communications Signals; 5, Service Communications Memory Device; 6, Long-Term Memory Device Reporting Situation in Network; 7, Operative Memory Device; 8, Scheduling Device; 9, Devite Receiving and Recording Signals from Test System; 10, Control Device; 11, Device Developing Interaction Signals; 12, Device Informing Ground Station of Situation in Network; 13, Device for Encoding Interaction Signals; 14, Service Communications Control Device

Figure 11.5 shows a block diagram of the control apparatus of the ground stations. The calls arriving over surface lines (number of called subscribers) move from input device 12 to the call reception and number recording device 11. The call reception device first determines the address of the subscriber called and determines, by referring to the memory unit for busy channels of the network 7 whether it is possible to provides a channel in the required direction. If the channels of the satellite or the ground station called are busy, this device outputs the busy signal to the subscriber; if there are free channels available on the satellite and at the called station, the subscriber is given a signal confirming acceptance of the call and the process of establishment of a connection begins.

First the call received is transmitted to the CC on instruction of the control device 9, for which it is sent to the device which forms service communications signals 10, then to coding device 3 and then to service communications transmitter 1 and the memory unit of the system for testing transmitted instructions 4, which checks the correctness of reception of messages from the GS at the CC.

The instruction of the CC for provision of a channel arrives through the service communications receiver 2 and is sent to the device for decoding of instructions 5 and then to instruction memory unit 6.

We should note that the instruction for provision of a channel is received simultaneously at the calling and called station and that all operations at these stations are performed similarly. At the same time the call is sent to the memory unit, it also arrives at the device signalling the servicing personnel to provides a channel 8. This is particularly important for the called station, but also necessary at the calling station. From the device which records the instruction, it is first sent to the CC for testing through the service communications transmitting /315
line -- coding device and transmitter. After reception of the signal indicating performance of the instruction, it is confirmed that the instruction was received properly, it is transmitted to the logic control device of the station, where the necessary control signals are developed for the receiving commutator 15 and transmitting commutator 14. These commutators supply the necessary frequencies from exciter 16 to the mixers of the channel transmitting device 17 and receiving device 18, connected to the proper surface line, thus providing a duplex satellite communi- /316
cations channel between two ground stations.

The device for testing connections made 13 and the CD of the station transmit a report of performance of instructions to the CC, and after reception of the next instruction from the CC, transmit the number of the called subscriber to the calling station, which number has been stored in the memory unit of the call receiving device. Transmission of the number signals that the channel is ready for operation.

§ 11.4. Control System for Centralized Network with Direct Establishment of Connection

The overall diagram of the operations of this network is similar to the diagram of Figure 11.3, relating to a network with centralized control. However, the operation of the CC and the control apparatus at the ground stations differs significantly. This is determined by the significantly greater independence of

the GS in the process of establishment of a connection and the correspondingly more modest role of the CC in this process.

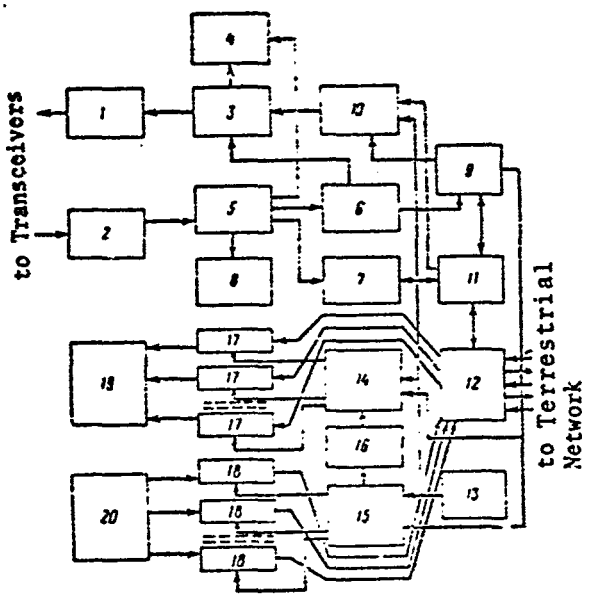


Figure 11.5. Control Apparatus at Low Channel Satellite Network Ground Station with Centralized Control: 1, Service Communications Transmitter; 2, Service Communications Receiver; 3, Device Coding Interaction Signals; 4, Device Recording Signals Transmitted; 5, Device Decoding Interaction Signals; 6, Device Recording Service Communications; 7, Long-Term Memory Carrying Network Load; 8, Device Signalling Arrival of a Call; 9, Control Device; 10, Device Froming Interaction Signals; 11, Device Receiving Interaction Signals from Surface Network and Memory Unit for Number of Called Subscribers; 12, Matching Device for Surface Communications Lines; 13, Device Testing Connections Made; 14, Transmitting Commutator; 15, Receiving Commutator; 16, Exciter; 17, Transmitting Channel Device; 18, Receiving Channel Device; 19, HF Transmitter; 20, HF Receiver

Figure 11.6 shows a functional diagram of a GS. Let us study the operation of the control system in servicing requests arriving at the GS from the surface network. A requirement, for example /317
a call, arrives over the surface line from a surface network switching center and is sent to input device 14, then to call

reception device 13, where the numbers of the called subscriber are stored until the connection is made.

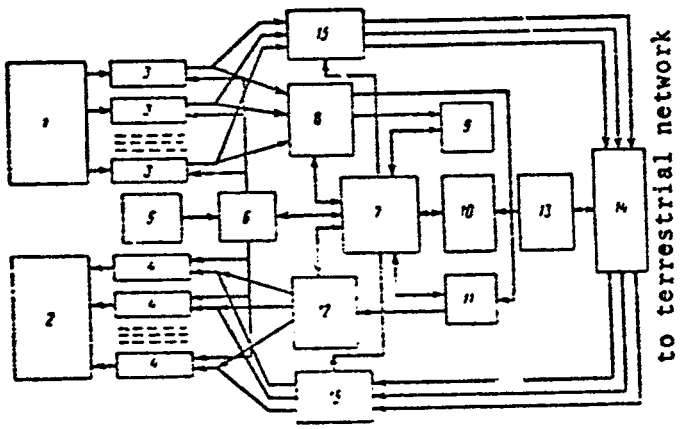


Figure 11.6. Functional Diagram of Ground Station of Low Channel Centralized Network with Direct Establishment of Connections: 1, HF Receiver; 2, HF Transmitter; 3, Receiving Channel Device; 4, Transmitting Channel Device; 5, Exciter; 6, Commutator for Signals from Exciter; 7, Control Device; 8, Device Testing Load on Network Channels; 9, Memory; 10, Device for Decoding Interaction and Address Signals; 11, Long Term Memory; 12, Device Producing Interaction Signals; 13, Device Receiving Interaction Signals from Surface Network and Recording Number of Called Subscriber; 14, Matching Device with Surface Communications Lines; 15, Channel Switching Device

Then the request is decoded 10, i.e., the type of request, address of the called subscriber if it is a call, etc. are determined. The decoded request is sent to control device 7. The request is also recorded in memory device 9 or long-term memory 11, depending on how rapidly it is to be processed. After recording of a request in memory, the devices for receiving and decoding of requests are released.

In order to provide a channel in a centralized network with direct establishment of connections, it is necessary to first find a free channel to the relay station, marked by the CC. Search for a free channel is performed on command of the CD by the device which tests the status of channels of the network 8, which continually observes all channels of the relay station.

After finding a free channel, the test device reports this to the control device, which sends instructions to commutator 6 to feed the required frequency from exciter 5 to the proper channel transmitter 4, and to the device which develops interaction signals 12 to send the busy signal to the CC. Removal of the marker and reception of a signal from the CC confirming that the channel is busy occur in the channel test device, which reports to the CD.

After receiving the busy confirmation signal, the CD gives an instruction to the commutators to adjust receiving channel 3 and transmitting channel 4 and an instruction to the interaction signal device to send the call signal to the called station. The response from the called station is also received by the channel testing device and sent to the CD, which connects the proper satellite channel to the surface line channel by means of channel switching device 15. After this, the CD instructs the device receiving requests to transmit the number of the called subscriber over the channel provided, which signals that the channel is ready for operation.

If the call arrives at the GS through the satellite from another station of the network, it is received by the channel test device, decoded and transmitted by it to the CD. The CD, according to the information contained in the call, sends an instruction to the commutator to feed the required frequency to the proper transmitting channnel device and an instruction to the interaction signal device to send a signal confirming reception of the call. Furthermore, the CD connects the satellite and surface channels.

A functional diagram of the control center of this network is presented in Figure 11.7.

Receiver device 1 of the CC simultaneously receives all satellite channels, more accurately those simplex channels of duplex pairs used for transmission from GS to the CS. Each of these channels is continually checked by the device for reception of interaction signals 3. /318

At the same time, the device developing interaction signals 4 marks all free channels of the satellite, i.e., transmits the marker signal in the simplex channel of each duplex pair designed for transmission from the CC to the CS. If any GS occupies a channel, the busy signal is received at the CC and sent to control device 5. The CD sends an instruction to the device developing interaction signals to remove the marker from the channel in question and send the confirming busy signal in this channel. If necessary, the CD can record the call from the GS in its main memory 6 or long-term memory 7.

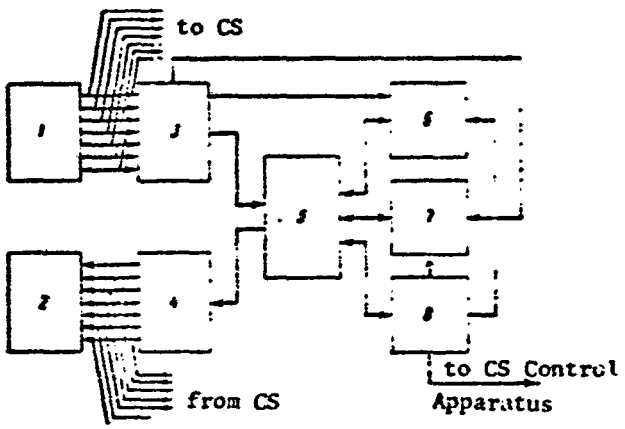


Figure 11.7. Functional Diagram of Control Center of Low Channel Network with Direct Establishment of Connections: 1, Receiver; 2, Transmitter; 3, Device Receiving Interaction Signals from Ground Stations; 4, Device Developing Interaction Signals; 5, Control Device; 6, Memory; 7, Long-Term Memory; 8, Device for Interaction with CS

The control center should be combined with the central station of the network, which generally has the heaviest traffic. In this case, the CC can use the transmitting and receiving equipment of the CS. The interaction of the CS with the CC is conducted by 8 in correspondence with the usual algorithm, but communications with them are of course not through the satellite /319 but directly through wires as shown on Figure 11.7.

After completion of communications, one of the GS, usually the calling station, sends the completion signal through the channel. This signal is received by the interaction signal receiver and is also transmitted after being decoded by the CD. The CD puts out and instruction to mark the now free channel.

§ 11.5. Control System for Low Channel Decentralized Network

In this network, the provision of channels is performed by ground stations, which must establish the connection without participation of the CC.

A functional diagram of the GS control apparatus is shown in Figure 11.8. The process of establishment of a connection begins with reception of a call from the surface network, arriving over lines connecting the GS to the nearest long-distance telephone

exchange or surface network switching center. Calls are fed through input device 17 to the unit which receives and decodes requests, 15. After decoding a call, this device reports the information on the call to the station CD 9 and, on instruction from the CD, records the number of the called subscriber in a special memory device 16, sends the address of the called station and other information concerning the call (category, priority, etc.) to memory 13 or long-term memory 14. /320

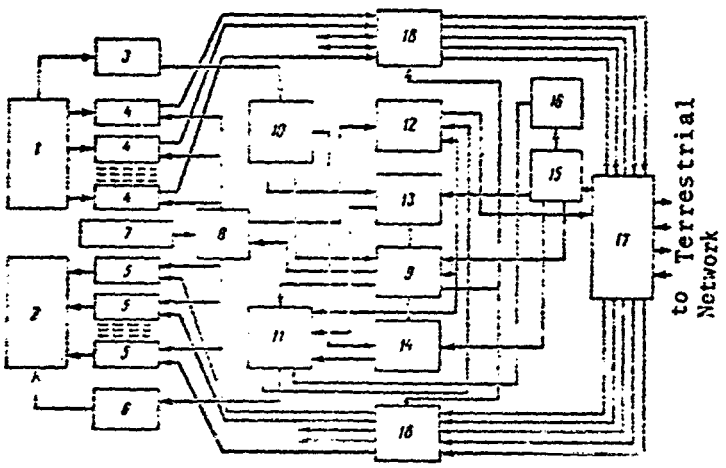


Figure 11.8. Functional Diagram of Ground Station and Low Channel Decentralized Communications Network: 1, HF Receiver; 2, HF Transmitter; 3, Service Communications Receiver; 4, Receiving Channel Device; 5, Transmitting Channel Device; 6, Service Communications Transmitter; 7, Exciter; 8, Exciter Signal Commutator; 9, Control Device; 10, Device for Reception of Interaction Signal; 11, Device for Transmission of Interaction Signal; 12, Device which Searches for Free Channels; 13, Memory; 14, Long-Term Memory; 15, Device for Reception and Decoding of Interaction Signals from Surface Network; 16, Device Recording Number of Called Subscribers; 17, Matching Device for Surface Lines; 18, Channel Switching Device

The control device then sends an instruction to the device developing interaction signals 11, to form a call to the GS with which communications must be established, using the information from memory, and transmit the call over the service communications line.

The message concerning the call includes the return address and channel number over which reception will be conducted at the calling station. This number is determined by the channel number over which the call was received at the GS from the surface network.

Subsequent operations involved in establishment of the connection are performed at the called station. Since the control apparatus at all GS is identical, let us study these operations using the same diagram, Figure 11.8.

The signal from the service communications receiver 3 from the calling GS is sent to the device for reception of interaction signals 10, where it is decoded. The call is then sent to the CD, and on its instruction is recorded in memory or long-term memory. Then the CD sends an instruction to the testing and channel search device 12 to find a free surface receiving and transmitting channel. After a free channel is found, the CD sends an instruction to commutator 8 to send the required frequencies from exciter 7 to the corresponding satellite receiver 4 and transmitter 5. When this occurs, a frequency is fed to the transmitting channel device according to the number of its receiving channel received from the calling GS, and to the receiving channel device, the frequency is sent which is determined by the CD.

After performing the necessary connections in the commutator, the CD, using the device which develops interaction signals 11, sends a signal to the called GS, confirming reception of the call and readiness for operation. Along with these signals, the number of the channel over which the called GS will receive the calling GS is transmitted. At the same time, the CD connects the satellite channels to the surface channels, sending the corresponding instructions to channel switching device 18.

The calling station receives the readiness signal and number of the receiving channel from the called station over the service line. These signals are sent to the device which receives interaction signals and then, in decoded form, to the CD. The CD sends an instruction to the commutator to connect the required exciter frequency to the transmitting channel device for transmission. The control device then connects the satellite channel to the surface channel and sends an instruction to transmit the called subscriber's number to the called station, which is the signal to begin operations. /321

At the end of the session, one of the subscribers (the first) sends the completion signal. This signal is sent to the GS and decoded in the device for reception of subscriber's requests. Then, the signal of completion is sent to the control device. The CD sends an instruction to the device forming interaction signals to output the "completion" signal to the other GS and then sends instructions to the commutator and channel switching device

to disconnect the circuits for transmission of exciter frequencies and the communications channel. At the same time, the information on a connection is erased in memory.

At the second GS, the operation involved when communications are completed are performed by the same method. However, the completion signal is received by the device for reception of interaction signals from GS and, furthermore, the CD sends no signal to the first GS, from which the completion signal has arrived. In order to increase the reliability of operation of service communications lines, as in the network with centralized control, it is possible to test messages transmitted by automatic return of the signals received and their comparison at the transmitting end. The apparatus required for operation of this type of test system is similar to that described above (see Figures 11.4 and 11.5) and therefore is not shown on Figure 11.8.

§ 11.6. Control System for Subtrunks in Multichannel Communications Networks

The primary distinction of multichannel communications is the operation in groups or bundles of channels, which defines the special structure of the control systems in these networks as well. In this case, the problem arises of switching groups of channels, i.e., subtrunks, as well as the problem of controlling channels within a subtrunk. In other words, the problem is that first each GS must be allocated a subtrunk for operations, then the channels of the subtrunk must be directed to the GS with which the station is to establish communications. Furthermore, the provision of duplex channel is required, i.e., the subtrunks radiated by called GS must include channels directed toward the calling GS and coupled with these channels, which this GS addresses to subscribers of the called GS.

Thus, without penetrating too deeply into the details of organization of communications as described earlier, we note only that in multichannel telephone satellite networks the process of control is a two-step process, in contrast to low channel networks.

The selection of a method of control of subtrunks, i.e., the first stage of control, is defined comparatively simply. The problem is that, as a rule, the number of subtrunks in any communications network cannot be too great, certainly not more than a few dozen, so that switching of subtrunks is not a frequent occurrence. Furthermore, it must be considered that the allocation of a multichannel subtrunk to any given GS has a significant influence on the operation of the communications network as a whole, in particular on the effectiveness of its operation and, consequently, the decision to allocate a subtrunk must be taken considering the situation over the entire network. These

/322

considerations lead us directly to the conclusion that in all networks (with any method of organization), control of subtrunks must be centralized and performed by the network control center. To achieve this control, the CC must continually receive information characterizing the status of the network and be able to evaluate the effectiveness of its operation with any distribution of subtrunks among stations.

Control of subtrunks is conducted from the control center according to the situation in the network, which is reflected in the memory devices of the CC. The basic indicators determining the situation are the existing distribution of subtrunks, the load on the subtrunks (with indication of direction of communications) and the streams of arriving requests at each station of the network.

Information on the distribution of subtrunks is received in the corresponding CC memory device from the central control device, since it is this CD which remotely controls all subtrunk commutators at GS of the network. Information on the loading of subtrunks, i.e., information on the number of operating channels available in each subtrunk and in which direction they are operating, is received at the CC from the ground stations over the service communications lines. The GS also send the CC information on the flow of requests arriving at the station. This information must be developed at the stations by a special device which records the incoming flow of requests, analyzes the corresponding factor characterizing the need of the station for channels[1].

Figure 11.9 shows that information from the GS arriving over service lines 1 after decoding 3 is sent by control device 6 to the corresponding memory unit -- the long-term station load memory unit 8 and the long-term GS requirements memory unit 9. Information, as was noted above, is sent to the subtrunk distribution memory 7 directly from the CD after reception of reports from GS subtrunk commutators that the instructions of the CD have been performed.

Thus, the three main memory units reflect the entire situation as concerns load of the network and the subtrunks required for control. The CC also has scheduling device 12, which contains information on necessary switching of subtrunks operating according to a schedule, memory unit 10 which assists the operation of the CD and situation analyzer 11. This analyzer /323

[1]We must note that at the present time the operating principles of such an automatic device is still not fully clear. Probably for the first period of operation an operator will be required to report the requirements of the station to the network.

continually analyzes the load on the network according to the information in memory and outputs recommendations concerning the optimal distribution of subtrunks.

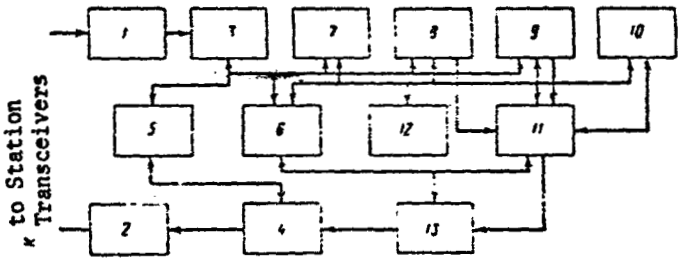


Figure 11.9. Functional Diagram of Subtrunk Control Center in Multichannel Satellite Network: 1, Service Communications Receiver; 2, Service Communications Transmitter; 3, Device for Decoding Interaction Signals; 4, Device for Coding Instructions; 5, Device for Testing Service Communications; 6, Control Device; 7, Subtrunk Distribution Memory; 8, Station Load Memory; 9, GS Requirements Memory; 10, Short-Term Memory; 11, Situation Analyzer; 12, Scheduling Device; 13, Instruction Output Device

The optimal distribution of subtrunks refers to the distribution of subtrunks among stations of the network such that the network is loaded to the maximum extent while the recommendations for servicing quality are fulfilled at all stations of the network. According to the recommendations of this device, the CD decides upon the allocation of a subtrunk to any given station and when one or more subtrunks should be removed from a station. Naturally, of course, before a subtrunk is removed from any GS, the subtrunk must be free. If several channels are busy in the subtrunk, attempts must be made to shift them to other subtrunks or to wait until they become free. The CC, observing the load of the channels at the GS, transmits information to all stations of the network concerning an overload at a station if it is determined that all channels at the station are busy. This information is required by the GS, as was stated above, to allow it to give a rapid busy signal if the called GS cannot respond to a call.

All decisions made by the CD are converted to messages in the instruction output device 13, then encoded 4 and sent to the service communications transmitter 2. In order to increase the reliability of operation of service communications lines, it is expedient to use the service communications test system 5. This system, as was described above, receives return signals and compares them with instructions transmitted. Only if the

instruction received back from a GS matches the instruction transmitted does the test system send the GS the instruction to perform the previous instruction. /324

A functional diagram of a subtrunk control apparatus as used at ground stations in the network is shown in Figure 11.10. The instructions to switch subtrunks are sent through the coding device 3 to GS control device 6. At the same time, these instructions may be transmitted to memory devices, separate for subtrunks received at the GS 7 and radiated by the GS 8, and the commutators of received subtrunks 10 and transmitted subtrunks 9, which connect the multiplexing apparatus (group devices) with the high frequency equipment at the outputs of the subtrunk receivers and inputs of the subtrunk transmitters.

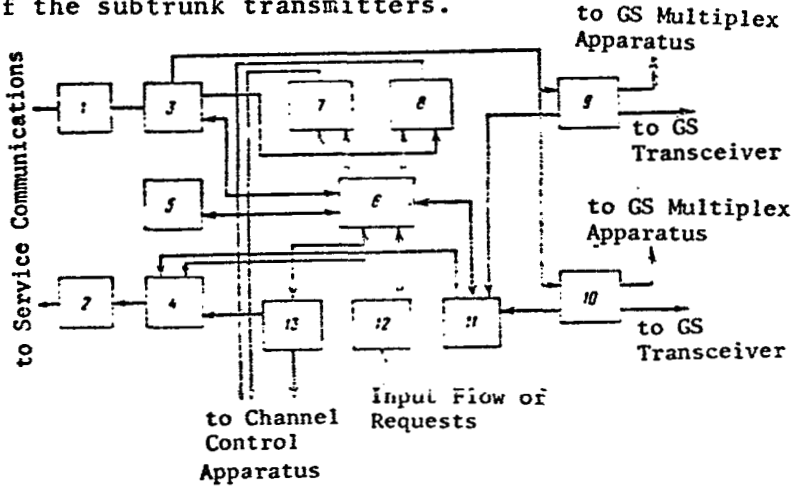


Figure 11.10. Functional Diagram of Subtrunk Control Apparatus at Network Ground Stations: 1, Service Communications Receiver; 2, Service Communications Transmitter; 3, Instruction Decoder; 4, Instruction Encoder; 5, Service Communications Test Device; 6, Control Device; 7, Memory Unit for Received Subtrunks; 8, Memory Unit for Transmitted Subtrunks; 9, Transmitted Subtrunk Commutator; 10, Received Subtrunk Commutator; 11, Device Testing Connections Made; 12, Device Analyzing Requests (Calls) Arriving; 13, Device Testing Channel Load

However, the sending of CC instructions to memory and to the commutators is controlled by the station CD. Only after the CC instruction is checked by the service communications test system 5 does the control device allow the device for decoding of instructions to send this instruction to the memory unit and commutators.

After the CC instruction is performed and the necessary connections are made in the commutators, test device 11 transmits a report on the connection made to the control center. The report from the test device is transmitted on instruction from the CD and /325 is sent first to encoder 4, then to service communications transmitter 2.

Information on the distribution of subtrunks recorded in memory is used in the operation of the channel control apparatus, which will be described below. It is sent to the control device of this apparatus. One extremely important function of the ground station control apparatus is collection and transmission of information on the load on the channels and the needs of the station for channels to the CC. In order to perform this function, the GS uses special load testing device 13 and incoming request analyzer 12.

The load testing device, according to the decisions of the CD, records the number of busy channels and the directions in which they are operating. Information on the number of operating channels in each direction is transmitted to the CC. This information may be transmitted either at certain predetermined time intervals or only when the situation concerning the load in some direction changes, i.e., the number of channels busy in some direction changes. Selection of the version of transmission of inforamtion on the load is determined by the operating mode of the station and the capabilities of the service communications lines.

At those times when the load on a station is great and the situation changes rapidly, it is expedient to use periodic transmission of information in order to avoid overloading the service communications lines. If the load on the station is slight or stable and the number of busy channels in each direction changes comparatively rarely, it is probably better to transmit information only at moments when the situation changes, for the same reasons.

The operation of the analysis device for incoming requests, as was noted above, cannot be precisely described at the present time. However, we can note that this device should operate integrally in principle, determining the magnitude of the incoming stream of requests and the number of calls satisfied for some time interval. Using these data, the device must decide on the requirements of the station for additional channels or the possibility of releasing some of the subtrunks already allocated to the station back to the control center.

§ 11.7. Channel Control System in Multichannel Communications Networks

Let us first study the operation of the apparatus controlling channels with the decentralized version of a control system. In

this case, all operations involved in establishment of connections are performed by the GS as they come in to communications, without participation of the control center (Figure 11.11).

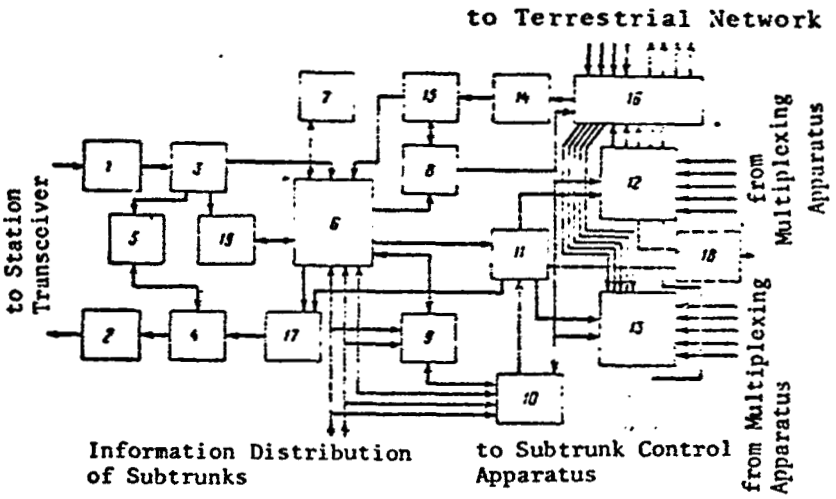


Figure 11.11. Functional Diagram of Ground Station Channel Control Apparatus in a Multichannel Network with Decentralized Control: 1, Service Communications Receiver; 2, Service Communications Transmitter; 3, Device for Decoding Interaction Signals; 4, Device for Encoding Interaction Signals; 5, Service Communications Test Device; 6, Control Device; 7, Operative Memory; 8, Called Subscriber Number Memory; 9, Device for Selection of Optimal Channel; 10, Device which Searches for Free Channels; 11, Marker; 12, Received Channel Commutator; 13, Transmitting Commutator; 14, Device for Reception of Interaction Signals from Surface Network; 15, Device for Decoding of Interaction Signals from Surface Network; 16, Matching Device with Surface Communications Lines; 17, Device for Formation of Interaction Signals; 18, Commutator Reforming Subtrunks in Network with Centralized Channel Control; 19, Network Station Busy Signal Memory Device

The call to a channel arrives in the form of a subscriber's number over a surface line connecting the surface network switching center to the satellite network ground station. The device for input of surface lines 16 is connected to the device for reception of interaction signals and calls 14. This device, receiving a call, sends it to decoding device 15, after which the /326

call is sent to the control section of the channel control device CCD[1] 6.

On command of the CCD, the number of the called subscriber is written in a special memory unit 8.

After receiving a call, the CCD first turns to the network station busy memory device 18 to see whether free channels are available to the station being called. If there are no free channels, the subscriber receives an immediate busy signal, while if there are, the process of establishment of a connection is begun. First the CCD gives an instruction to the devices which search for a free channel 10 and select the optimal channel 9, which operate together closely to find a satellite channel for transmission of /327
information to the GS called. During the process of search for a free channel, these devices use information on the distribution of subtrunks recorded in the memory of the subtrunk control apparatus of the GS.

A free channel is selected so as to ensure the most effective operation of the stations. In particular, free channels are first sought in the most heavily loaded subtrunk, so that in case the load should drop the more lightly loaded subtrunks will be more rapidly completely freed and can be returned to the CC. Within each subtrunk, free channels not included in a triplet[2], free primary groups, etc. are used first. This is done so that if necessary a triple channel or wide band channel occupying the bandwidth of a primary group can be organized without additional switching.

After finding the free channel, the search and selection devices report their decision to the CCD, which sends an instruction to marker 11 to make the required switching in transmitting commutator 13. At the same time, the CCD sends a call to the called GS and the information as to which channel the calling GS. will use to transmit its information to the called GS.

We will now analyze the process of establishment of a connection, using the same diagram of Figure 11.11.

Calls from the ground station of the network arrive through the service communications terminal devices 1 and 3 at the channel control device. The CCD gives an instruction, as at the calling

---

[1]We should note that the CCD can be combined with the device for control of GS subtrunks (see Figure 11.10). In this same way, the service communications devices can also be combined.

[2]Meaning a triple-width channel designed for transmission of radio broadcases or other wide band information.

GS, to the devices for search and selection of channels, to find a free surface duplex channel from the station and a satellite channel over which the GS will answer the call.

After producing the decision on selection of channels, the CCD gives an instruction to the marker to make the necessary connections in the commutator and at the same time sends a signal to the calling GS to confirm reception of the call and information on the channel over which the GS will respond. The calling GS, after receiving the answer from the called GS, uses the same algorithm to make its connection in receiving commutator 12. After this, the channel can be considered ready for operation. The device for control of channels in the calling station gives an instruction for transmission of the number of the called subscriber over this channel as a signal to begin operations over service communications lines 4 and 2.

After completion of a conversation, the completion signal, given by one of the subscribers (the first) is sent through the interaction signal receiver to the CCD. Receiving the signal, the CCD, using the marker, determines the numbers of the satellite channels over which the information was transmitted and received, then gives instruction to the marker to break the connection. Together with this, the CCD of the GS which first received the completion signal sends the corresponding information to the GS in correspondence with it. This information reports the number of the simplex satellite channels which formed the duplex channel no longer operating. The CCD of the second GS, after receiving and decoding this information, also disconnects the lines in its commutators. /328

The operation of the channel control apparatus at ground stations with centralized construction of the system is quite similar to the operation with decentralized construction.

At the peripheral stations of the network, the functional plan of the channel control apparatus does not differ from the plan shown in Figure 11.11. True, the volume of this apparatus will be significantly less, particularly the volume of the receiving commutator and marker, since the PS receives a limited number of subtrunks only from the CS. The order of operation of the apparatus in the process of establishment of a connection is also similar to the operation of the GS described above.

The central station of the network differs from that described above basically in the commutator as well, which must service not only PS-CS communications, the process of establishment of connections in which is similar to that described, but also communications of peripheral stations with each other.

For PS-PS communications, the central station of the network carries a third commutator 18 (Figure 11.11), designed for reforming of PS subtrunks. This commutator is installed after the receiving commutator, receiving all its information from the satellite, receives only those channels which are directed toward the PS. The commutator forms groups directed toward peripheral stations, then these groups are sent to the proper sets of multiplexing apparatus and transmitters.

Naturally, the presence of a third high-capacity commutator[1] sharply complicates both the control device and most particularly the marker. The operation of the free channel search device is also more complex, as well as that of the device for selection of the optimal channel, lobe testing system and channel distribution system and the system testing loading and distribution of subtrunks. However, the functions performed by these devices do not differ from the functions of similar devices in Figure 11.11.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1974—535-042/7

[1]It may be combined with the receiving commutator.

END

DATE

FILMED

DEC 9 1974